

**PCT**

WELTORGANISATION FÜR GEISTIGES EIGENTUM  
Internationales Büro



INTERNATIONALE ANMELDUNG VERÖFFENTLICHT NACH DEM VERTRAG ÜBER DIE  
INTERNATIONALE ZUSAMMENARBEIT AUF DEM GEBIET DES PATENTWESENS (PCT)

<b>(51) Internationale Patentklassifikation</b> <sup>6</sup> : <b>C07H 21/00, C07K 1/04, C12N 15/10, C12P 21/00</b>		<b>A1</b>	<b>(11) Internationale Veröffentlichungsnummer:</b> <b>WO 95/17413</b> <b>(43) Internationales Veröffentlichungsdatum:</b> <b>29. Juni 1995 (29.06.95)</b>
<b>(21) Internationales Aktenzeichen:</b> <b>PCT/EP94/04240</b> <b>(22) Internationales Anmeldedatum:</b> <b>20. December 1994 (20.12.94)</b>		<b>(81) Bestimmungsstaaten:</b> JP, US, europäisches Patent (AT, BE, CH, DE, DK, ES, FR, GB, GR, IE, IT, LU, MC, NL, PT, SE).  <b>Veröffentlicht</b> <i>Mit internationalem Recherchenbericht.</i>	
<b>(30) Prioritätsdaten:</b> <b>P 43 43 591.2 21. December 1993 (21.12.93) DE</b>			
<b>(71) Anmelder (für alle Bestimmungsstaaten ausser US):</b> EVOTEC BIOSYSTEMS GMBH [DE/DE]; Grandweg 64, D-22529 Hamburg (DE).			
<b>(72) Erfinder; und</b> <b>(75) Erfinder/Anmelder (nur für US):</b> SCHWIENHORST, Andreas [DE/DE]; Weender Landstrasse 68, D-37075 Göttingen (DE). LINDEMANN, Björn [DE/DE]; Tannenhof 115b, D-22397 Hamburg (DE). FUCHS, Merle [DE/DE]; Stiegel 5, D-37077 Göttingen (DE). HENCO, Karsten [DE/DE]; Kirchberg 4, D-40699 Erkrath (DE).			
<b>(74) Anwälte:</b> MEYERS, Hans-Wilhelm usw.; Deichmannhaus am Hauptbahnhof, D-50667 Köln (DE).			
<b>(54) Title:</b> PROCESS FOR THE EVOLUTIVE DESIGN AND SYNTHESIS OF FUNCTIONAL POLYMERS BASED ON DESIGNER ELEMENTS AND CODES			
<b>(54) Bezeichnung:</b> VERFAHREN ZUM EVOLUTIVEN DESIGN UND SYNTHESE FUNKTIONALER POLYMERE AUF DER BASIS VON FORMENELEMENTEN UND FORMENCODES			
<b>(57) Abstract</b>  A process for the production of oligomeric or polymeric functional elements from designer elements in which the functional elements are obtainable by the linking of at least two designer elements, at least one of which is itself made up of at least two monomers which are linked by at least one chemical bond which corresponds to the chemical bond between two designer elements.			
<b>(57) Zusammenfassung</b>  Verfahren zur Herstellung oligomerer oder polymerer Funktionselemente aus Formenelementen, wobei die Funktionselemente erhältlich sind durch Verknüpfung von mindestens zwei Formenelementen, von denen mindestens ein Formenelement selbst aus mindestens zwei Monomeren aufgebaut ist, die durch mindestens eine chemische Bindung verknüpft sind, die der chemischen Bindung zwischen zwei Formenelementen entspricht.			

#### ***LEDIGLICH ZUR INFORMATION***

**Codes zur Identifizierung von PCT-Vertragsstaaten auf den Kopfbögen der Schriften, die internationale Anmeldungen gemäss dem PCT veröffentlichen.**

<b>AT</b>	Österreich	<b>GA</b>	Gabon	<b>MR</b>	Mauretanien
<b>AU</b>	Australien	<b>GB</b>	Vereinigtes Königreich	<b>MW</b>	Malawi
<b>BB</b>	Barbados	<b>GE</b>	Georgien	<b>NE</b>	Niger
<b>BE</b>	Belgien	<b>GN</b>	Guinea	<b>NL</b>	Niederlande
<b>BF</b>	Burkina Faso	<b>GR</b>	Griechenland	<b>NO</b>	Norwegen
<b>BG</b>	Bulgarien	<b>HU</b>	Ungarn	<b>NZ</b>	Neuseeland
<b>BJ</b>	Benin	<b>IE</b>	Irland	<b>PL</b>	Polen
<b>BR</b>	Brasilien	<b>IT</b>	Italien	<b>PT</b>	Portugal
<b>BY</b>	Belarus	<b>JP</b>	Japan	<b>RO</b>	Rumänien
<b>CA</b>	Kanada	<b>KE</b>	Kenya	<b>RU</b>	Russische Föderation
<b>CF</b>	Zentrale Afrikanische Republik	<b>KG</b>	Kirgisistan	<b>SD</b>	Sudan
<b>CG</b>	Kongo	<b>KP</b>	Demokratische Volksrepublik Korea	<b>SE</b>	Schweden
<b>CH</b>	Schweiz	<b>KR</b>	Republik Korea	<b>SI</b>	Slowenien
<b>CI</b>	Côte d'Ivoire	<b>KZ</b>	Kasachstan	<b>SK</b>	Slowakei
<b>CM</b>	Kamerun	<b>LI</b>	Liechtenstein	<b>SN</b>	Senegal
<b>CN</b>	China	<b>LK</b>	Sri Lanka	<b>TD</b>	Tschad
<b>CS</b>	Tschechoslowakei	<b>LU</b>	Luxemburg	<b>TG</b>	Togo
<b>CZ</b>	Tschechische Republik	<b>LV</b>	Lettland	<b>TJ</b>	Tadschikistan
<b>DE</b>	Deutschland	<b>MC</b>	Monaco	<b>TT</b>	Trinidad und Tobago
<b>DK</b>	Dänemark	<b>MD</b>	Republik Moldau	<b>UA</b>	Ukraine
<b>ES</b>	Spanien	<b>MG</b>	Madagaskar	<b>US</b>	Vereinigte Staaten von Amerika
<b>FI</b>	Finnland	<b>ML</b>	Mali	<b>UZ</b>	Usbekistan
<b>FR</b>	Frankreich	<b>MN</b>	Mongolei	<b>VN</b>	Vietnam

Verfahren zum evolutiven Design und Synthese funktionaler Polymere auf der Basis von Formenelementen und Formencodes

Gegenstand der vorliegenden Erfindung ist ein Verfahren gemäß Anspruch 1.

Die rasante Entwicklung der letzten Jahre im Bereich der Biowissenschaften hat nicht nur die Grundlagenforschung, sondern gerade auch die angewandte Forschung in diesem Feld stimuliert. Proteine spielen hier aufgrund ihres breiten Wirkungsspektrums eine herausragende Rolle. Ein ganzer Zweig der modernen Biotechnologie beschäftigt sich heute mit dem sog. "Protein Engineering", d.h. der Herstellung von Designer-Proteinen, die entweder auf der Grundlage bekannter Proteine durch graduelles Abändern oder durch vollständige Neusynthese entwickelt werden. Man unterscheidet hier vor allem zwei Ansätze, das rationale und das irrationale Design.

Rationales Design ist darauf aus, eine Aminosäuresequenz zu produzieren, die sich in eine gewünschte Struktur faltet, und zusätzlich die erhoffte Funktion aufweist. Damit hängt diese Strategie ganz offensichtlich von einem tiefen Verständnis des "Protein folding" ab. Fortschritte in den letzten Jahren betrafen u.a. das rationale Design einfacher

Struktur-Domänen. Das Design größerer Proteine mit komplexen oder gar beispiellos neuen Eigenschaften liegt jedoch immer noch außerhalb der Möglichkeiten dieses Ansatzes. Demgegenüber setzt irrationales Design keine Informationen über die Proteinstruktur, Proteinfaltung etc. voraus. Einzig die Kenntnis der gewünschten Eigenschaft und eine Möglichkeit, Molekülpopulationen gemessen an dieser Eigenschaft zu bewerten, sind hier Voraussetzung. Ausgehend von einer "combinatorial library" aus Peptiden oder Proteinen werden Moleküle mit der gewünschten Eigenschaft selektiert und erst im Nachhinein analysiert. Hier wird also der Mechanismus, nachdem ein Molekül die gestellte Aufgabe meistert, nicht im voraus determiniert.

Obwohl dieser Ansatz in sehr eleganter Weise gerade auch in jüngster Zeit Peptide mit einfachen und z. T. neuen Eigenschaften hervorgebracht hat, stellt sich auch hier das Problem, wie man zu größeren Proteinen mit komplexeren Funktionen kommen kann. Schon eine vollständige Bank eines 20mers liefert mit  $20^{20} = 10^{26}$  verschiedenen Sequenzen eine astronomisch hohe Zahl zu untersuchender Moleküle. Soll die Peptidsequenz auch noch durch eine Nucleinsäure codiert werden, stellt sich das Problem in noch gravierender Weise. Da der genetische Code degeneriert ist, d.h. eine Aminosäure u.U. durch mehrere verschiedene Codons repräsentiert wird, ergibt sich hier eine Zahl von mindestens  $4^{60} = 10^{36}$  Molekülen, die synthetisiert werden. Normalerweise wird an der dritten Codonposition nur G oder C zugelassen, um Stopcodons weitgehend zu vermeiden. Die verbleibende Zahl von  $10^{30}$  Molekülen übersteigt noch immer die Standardausbeute einer kommerziellen DNA-Synthese um 12 Größenordnungen. Eine weitere Reduktion der pro Position zugelassenen Codons wurde von Youvan vorgeschlagen. Ob diese Methode den meßbaren Sequenzraum nicht in unzulänglicher Weise einschränkt, gerade bei der Suche nach neuen Funktionen, bleibt abzuwarten.

- 3 -

Zum Aufbau funktionaler Strukturen arbeitet die Natur mit modularen Systemen. Bekannt sind die Nucleotidbausteine, die Aminosäure-Bausteine (als Nucleotidtriplets kodiert) und Exon-Domänen (aus Aminosäurebausteinen aufgebaut). Die evolutive Optimierung funktionaler Biopolymere entsprechend der Patentanmeldung WO 92/18645 geht von der Vorstellung aus, durch kontinuierliche Verbesserung bestehender Basiseigenschaften, z. B. eines Enzyms, bei der kontinuierlichen Anpassung an erwünschte Reaktionsbedingungen wie Ionenstärke, Temperatur, pH-Wert eine optimale Struktur zu finden. Sind vorteilhafte oder mindestens neutrale Mutationen möglich, so sind durch mehrmalige Wiederholungen von Selektion und Mutation auch entfernte Bereiche des Sequenzraumes zugänglich, die durch die Ausgangspopulation nicht abgedeckt waren. Von der ursprünglichen, bereits funktionsfähigen Struktur entfernt man sich jedoch bei diesem Vorgehen in keinem Schritt. Optimiert wird eine Eigenschaft des Ausgangsmoleküls, die bereits - wenn auch in bescheidenem Maße - im ursprünglichen Molekül inhärent ist. Der "Pfad", den eine solche Evolution durch den Sequenzraum nimmt, ist bestimmt durch die zugänglichen, in Richtung der Optima führenden Grate in der unterliegenden Wertelandschaft. Wie bei allen Methoden, die den Sequenzraum nicht vollständig erschließen, besteht bei diesem Vorgehen die nur schwer einzuschätzende Gefahr, in einem lokalen Optimum stecken-zubleiben. Für die Praxis bedeutet dies, daß bestimmte Regionen des Sequenzraumes einschließlich der dort befindlichen Optima, durch breite und tiefe "Täler" abgetrennt sind. Bei der begrenzten Populationsgröße von Molekülspezies in Experimenten (P 43 22 147, WO 92/18645) ist aber die Wahrscheinlichkeit zu niedrig, entfernte Vielfehlermutanten zu erzeugen, die sich jenseits dieser Schranke befinden und den Weg zu diesen neuen Optima anzeigen.

Die Natur hat eine Anzahl von Mechanismen entwickelt, mit dieser Problematik umzugehen: lange Entwicklungszeiträume, Rekombinationsverfahren (horizontaler Gentransfer, Crossing-

over, Genkonversion, Exon-Rekombination (exon-shuffling), Virusshuttles, mobile Elemente (Transposons), Untereinheiten-Struktur von komplexen Proteinen)) sowie Multigenfamilien mit Pseudogenen.

Mit der Anzahl der parallel geführten Mutantenbildung und Selektion lässt sich die Chance auf Erzeugung einer gewünschten Vielfehlermutante erhöhen; durch Rekombination lassen sich mutierte Gensegmente effizient mischen. Funktionslose Pseudogene als Mitglieder einer funktionsfähigen Multigenfamilie lassen sich als Vielfehlermutanten auch über längere Entwicklungszeiträume ohne Gegenselektion in ihrer Existenz erhalten, um eventuell bei Rückerhaltung einer Funktion wieder positiv selektierbar zu werden.

Die Übertragung dieser Mechanismen auf eine effiziente *in vitro* Optimierung ist offensichtlich nicht ohne weiteres möglich. Die Schwierigkeiten müssen jedoch in jedem Falle für solche Aufgabenstellungen gelöst werden, bei denen eine kontinuierliche Optimierung nicht erwartet werden kann. Dies trifft insbesondere für solche Anpassungsprozesse zu, bei denen eine Funktion vollständig neu etabliert werden muß.

Das der Erfindung zugrundeliegende technische Problem betrifft die Bereitstellung eines Verfahrens zur Herstellung oligomerer oder polymerer Funktionselemente wie Biopolymere mit funktionalen Eigenschaften, beispielsweise Enzymen, Ribozymen, Wirkstoffen, etc. Dabei soll unter Ausnutzung evolutiver Strategien ein den herkömmlichen Screening-Verfahren überlegenes Verfahren bereitgestellt werden.

Gelöst wird dieses Problem durch ein Verfahren mit den Merkmalen des Anspruchs 1. Die sich daran anschließenden Unteransprüche betreffen bevorzugte Ausführungsformen des erfindungsgemäßen Verfahrens.

- 5 -

Erfindungsgemäß werden zur Herstellung oligomerer oder polymerer Funktionselemente aus Formenelementen zunächst Formenelemente durch chemische oder enzymatische Verknüpfung von mindestens zwei Monomeren aufgebaut und die so erhältlichen Formenelemente dann zu Funktionselementen verknüpft. Dabei entspricht die Natur der chemischen Bindung zwischen den Monomeren derjenigen zwischen den jeweiligen Formenelementen. Die so erhältlichen Funktionselemente können dann auf die bestimmten potentiellen Funktionen getestet werden. Die Vorteile der erfundungsgemäßen Vorgehensweise werden durch die nachfolgende Beschreibung weiter verdeutlicht.

Bevorzugt wird die Verknüpfung der Formenelemente unter Einsatz einer festen Phase als Reaktionsträger durchgeführt. Die Verknüpfung der Formenelemente kann chemisch und/oder enzymatisch erfolgen. Die Verknüpfung der Formenelemente zu den Funktionselementen kann entweder planmäßig durch gezielte Zugabe der einzelnen Formenelemente und nachfolgender Verknüpfung oder auch statistisch durch zufällig gesteuerte Zugabe der Funktionselemente und deren Verknüpfung erfolgen. Es ist dabei möglich, die Verknüpfung schrittweise aufbauend stereospezifisch und/oder gerichtet durchzuführen.

Als Formenelemente kommen vorzugsweise Nucleinsäuren, doppelsträngige oder einzelsträngige DNA und/oder RNA und/oder modifizierte Nucleinsäuren in Frage. Als Formenelemente kommen auch Peptide und/oder Polypeptide und/oder sonstige kopplungsfähige chemische Oligomer-Formenelemente in Frage. Dazu können auch Oligo- oder Polysaccharide gehören.

In einer bevorzugten Ausführungsform des erfundungsgemäßen Verfahrens werden die Formenelemente als bereits synthetisierte Oligomerbausteine eingesetzt oder im Reaktionsgefäß quasi in situ hergestellt.

Es ist vorteilhaft, die Reaktion der Formenelemente in parallel geführten Mikroreaktionsansätzen (wie in P 43 22

147.5 vorgeschlagen) durchzuführen, bei denen die Formenelemente in vorbestimmter Reihenfolge verknüpft werden. Insbesondere werden nach erfolgter Synthese die Reaktionsprodukte wie Funktionselemente oder Vorstufen davon an der festen Phase gebunden bleiben und nach Abtrennung der Reaktionspartner weiter verarbeitet oder von der Festphase entkoppelt. Es ist jedoch ebenfalls möglich, die Reaktion in geeigneten, dem Fachmann bekannten Reaktionsbedingungen in Lösung durchzuführen oder die festphasengekoppelte oder in homogener Lösung durchgeführte Reaktion miteinander zu kombinieren.

Durch Einsatz der Fluoreszenzkorrelationsspektroskopie (FCS) (PCT/EP 93/01291) wird es ermöglicht, die Funktionsweise der Funktionselemente im gleichen Volumenelement direkt zu bewerten, in dem auch die Synthese abläuft. Dies bedeutet eine sehr direkte Möglichkeit, das Ergebnis einer aufbauenden Funktionselementssynthese zu kontrollieren.

Vorzugsweise wird pro Reaktionsschritt, bei der schrittweisen Verknüpfung der Formenelement, jeweils ein Formenelement als Reaktionspartner an fester Phase gekoppelt. Es können auch Mischungen von Formenelementen eingesetzt werden und/oder im Reaktionsgefäß direkt generiert werden. Werden als Formenelemente Nucleinsäuren verwendet, so ist es vorteilhaft, wenigstens einen Reaktionspartner mit einer Schnittstelle eines Restriktionsenzym zu versehen oder ein Nucleinsäure-formenelement zu verwenden, welches frei von Start- und/oder Stopcodons ist. Vorzugsweise sind die Reaktionsschnittstellen solche, die von Restriktionsenzymen der Klasse IIS erkannt werden können. Die Einführung von Restriktionsschnittstellen dieser Enzymklasse ist vorteilhaft, da beliebige Sequenzen gerichtet verknüpft werden können, ohne daß die Wahl des Reaktionssenzyms die Sequenzerfordernisse des Endproduktes beeinflußt.

- 7 -

Sind in den zu verknüpfenden Formenelementen einzelsträngige Überhänge eingeführt, so können darüber beliebige Sequenzen gerichtet verknüpft werden, ohne daß dabei irgendwelche Anforderungen an die Sequenz des gewünschten Endproduktes gestellt werden müssen. Dieses Erfordernis kann auch durch selektive und reversible chemische und/oder enzymatische Modifikation der 3'- und/oder 5'-Enden der Nucleinsäuren, zum Beispiel durch Phosphorylierung anstelle und in Kombination mit der Einführung der einzelsträngigen Überhänge erzielt werden.

Als Beispiel für eine reversible chemische Modifikation ist das Ankoppeln einer Trityl-Schutzgruppe, die durch Behandlung mit Essigsäure abspaltbar ist, zu nennen. Die Einführung der Tritylgruppe am 3'- oder 5'-Ende des Nucleotids führt zur Blockade der Ligation der Formen- und/oder Funktionscodes bzw. -elemente.

Durch Behandlung eines Oligo- oder Polynucleotids mit Nuclease kann ein 3'- oder 5'-Ende modifiziert werden, zum Beispiel wird durch Behandlung mit Exonuclease III das 3'-Ende durch Abdauung modifiziert. Wenn in das entsprechende Oligo- oder Polynucleotid (z. B. DNA) Nucleotidtriphosphate eingebaut werden, so stoppt die Exonuclease am ersten Thio-Nucleotid die Abdauung. Damit ergibt sich eine regulierbare Modifizierung des Endes des Oligo- oder Polynucleotids.

Das erfindungsgemäße Verfahren erlaubt den Einsatz von Formenelementen, die nach röntgen-kristallographisch analysierten natürlichen Funktionsdomänen von Proteinen und Polypeptiden bekannt sind. Es können so bereits bekannte Bausteine bzw. Module von in der Natur bereits vorkommenden Funktionselementen benutzt werden.

Die zu verwendenden Formenelemente können auch aus Selektionsexperimenten gewonnen werden.

Insbesondere vorteilhaft ist die Verwendung von Formenelementen in einer Länge von 1 bis 60 Aminosäuren oder Nucleotidsequenzen entsprechender Kodierungslänge. Die Formenelemente können auch an bestimmten Positionen degeneriert sein und/oder Deletionen oder Insertionen tragen, insbesondere bei Verwendung von Nucleotiden als Formenelemente.

Es wird auch die Verwendung des erfindungsgemäßen Verfahrens wie oben beschrieben zur Synthese parallel aufgebauter Formen-Bibliotheken funktionaler Oligomere oder Polymere beansprucht.

Die ursprüngliche Aufgabe von "combinatorial libraries" ist eher das Angebot einer Funktionen-Vielfalt als einer Sequenzvielfalt. Es ist heute eine Tatsache, daß die dreidimensionale Struktur von Proteinen relativ stabil gegen Substitutionen einzelner Aminosäuren ist. Durch die große Zahl aufgeklärter Proteinstrukturen gewann man die Erkenntnis, daß Proteine zwar keine oder nur sehr geringe Sequenzhomologie aufweisen können, aber trotzdem die gleiche oder sehr ähnliche 3D-Struktur einnehmen können. Dies beruht möglicherweise darauf, daß nur eine begrenzte Anzahl möglicher Faltungsweisen von Aminosäureketten unter biologischen Bedingungen stabil ist. Strukturelle Verwandtschaft spiegelt aber auch die Evolution rezenter Proteine aus einer relativ begrenzten Zahl von Ur-Strukturen, -Modulen heraus wieder. Diese Module können als kleine, funktionelle Domänen oder kompakte Struktureinheiten verstanden werden und können auch in heutigen Genen leicht aufgespürt werden. In der Hypothese des "Exon-shuffling" wird vermutet, daß die Evolution zu komplexeren Proteinen gerade durch die Kombination von Exons, also Modulen im oben beschriebenen Sinn enorm beschleunigt wurde. Wenn man annimmt, daß die Zahl der Exons, die die Konstruktion aller heute bekannten Proteine erlauben würde, zwischen 1000 und 7000 zu suchen ist, eröffnet eine hierarchische Strategie des "Protein Design" mit Bausteinen zunehmender Komplexität die Möglichkeit der viel schnelleren

Durchmessung eines "shape space" mit zugehörigerer "fitness landscape" als es die Suche in einer traditionellen "combinatorial library" gestatten würde. Ein Protein aus 150 Aminosäuren (die Größe einer klassischen Nucleotid-bindungsstelle, der sog. "Rossman fold") müßte nach herkömmlichem Verfahren aus einer Bibliothek von  $20^{150} = 10^{195}$  verschiedenen Aminosäuresequenzen selektiert werden. Kombinationen von 1000 verschiedenen Modulen der Länge 30 Aminosäuren ergeben hingegen nur eine Komplexität von  $1000^5 = 10^{15}$  Molekülen.

Das erfindungsgemäße Verfahren ist ein hierarchisches Verfahren zum Design von Proteinen, Nucleinsäuren deren Derivaten oder chemischer Oligo- oder Polymere mit bestimmten gewünschten Eigenschaften, ausgehend von Modul-Bibliotheken, im folgenden als Formenelemente bezeichnet. Erfindungsgemäß können die Formenelemente auch Gensegmente sein, die für Formenelemente kodieren. Die als Module fungierenden Formenelemente sollen zufällig kombinierbar sein. Kleinere Proteine oder Untereinheiten für größere Proteine mit bestimmten Eigenschaften werden in einem anschließenden Selektionsschritt aus dem Pool von Modulkombinationen herausgesondert und können ihrerseits wieder als Bausteine in einer Untereinheits-Bibliothek dienen, u.s.w.

Auf jeder Konstruktionsstufe kann durch fehlerhafte Kopierung einzelner Bausteine zusätzlich ein "Rauschen" auf Aminosäuresequenzebene eingeführt werden. Dies ermöglicht die Modularierung der dreidimensionalen Anordnung chemischer Gruppen und somit eine weitere funktionelle Optimierung selektierter Moleküle. Die vorgeschlagene Strategie erfordert eine neue Art von "Artificial Gene Assembly".

Bisher werden vor allem zwei Methoden angewandt, denen gemeinsam ist, daß die DNA in einer bestimmten Orientierung ligiert wird, um damit auch die Abfolge der Aminosäuren festzulegen. Die wohl älteste Methode - von Khorana und

- 10 -

seinen Mitarbeitern entwickelt - arbeitet mit überlappend komplementären einzelsträngigen DNA Molekülen, die vor der Ligation miteinander hybridisiert werden. Die zweite Methode nutzt Schnittstellen von Restriktionsenzymen im zu konstruierenden Gen, um an diesen Stellen das Gen in Blöcke zu unterteilen, die dann in mehreren aufeinanderfolgenden Schritten zusammengesetzt werden. Durch beide Methoden wird die Sequenz an den Übergängen der verwendeten Oligo-DNAs bzw. Blöcke methodisch bedingt festgelegt. Dies aber entspricht gerade nicht der Anforderung nach beliebiger Austauschbarkeit der einzelnen Module schon in der Konstruktionsphase des Gens. Bestandteil der vorliegenden Erfindung ist also notwendigerweise auch eine neue Art des "Artificial Gene Assembly". Erfindungsgemäß wird in allgemeiner Form wie folgt verfahren:

- Das Verfahren des "Artificial Gene Assembly" arbeitet analog des in der WO 92/18645 beschriebenen Verfahrens;
- das Verfahren erschließt nicht den Umgang mit der Varianz im Sequenzraum sondern mit der Varianz im sogenannten Formenraum. Der Formenraum, gebildet aus Basiselementen definierter stabiler Formenelemente, reduziert die Komplexität der Varianten der Bauelemente des Sequenzraumes;
- das Verfahren erschließt den Funktionsraum über eine Variation von Bausteinen des Formenraumes;
- als Bausteine werden Bausteine des Form-Codes (siehe unten) eingesetzt;
- für die Auswahl der Bausteine werden bestimmte Auswahlkriterien zur Vorselektion eingesetzt, die theoretischen Annahmen entsprechen oder natürlichen Formen-Analoga entsprechen.

- 11 -

Als Module zur parallel geführten Variation (Mutation) und Selektion stehen bislang nur die Nucleotide oder Aminosäuren als synthetisch oder enzymatisch handhabbare Bausteine eines Polymers für gerichtete Kopplungsprozesse zur Verfügung. Der direkte Zugang zu einer funktionalen Oberflächenstruktur eines Polymers scheitert wie oben angeführt in vielen Fällen am Problem der großen Zahlen der Varianten des Sequenzraumes.

Gegenstand dieses evolutiven Anpassungsprozesses ist der Einsatz modularer Bausteine, dem Formencode, bestehend aus den Formenelementen. Der Formencode umfaßt Formenelemente, aufgebaut aus Elementen des Sequenzraumes. Der Formencode, wie er beispielsweise aus natürlichen Polymeren wie Proteinen, Polypeptiden oder funktionalen Nucleinsäuren abgeleitet werden kann, kodiert unter festgesetzten äußeren Bedingungen stabile Formenelemente (Sekundärstrukturen, eventuell Tertiärstrukturelemente enthaltend). Dabei ist bemerkenswert, daß sehr unterschiedliche Sequenzen (Primärstrukturen) für sehr ähnliche Formenelemente kodieren können. Mit anderen Worten, im Formenraum sehr eng benachbarte Elemente können im Sequenzraum sehr weit voneinander entfernt liegen (große Hamming Distanz). Das gleiche gilt für den umgekehrten Fall. Im erfindungsgemäßen Sinne erklärt eben diese Eigenschaft, daß bereits der Austausch formenmäßig gleicher Sequenzen im Sequenzraum einen großen Schritt im Sinne einer Vielfehlermutante bedeuten kann. Mit Hilfe der erfindungsgemäßen angesprochenen Syntheseverfahren ist diese Anforderung technisch umsetzbar. Die Erstellung der entsprechenden Verteilungen gelingt durch programmierte Synthese. Sie ist nicht, wie in WO 92/18645 beschrieben, durch fehlerhafte Replikation im Sinne fehlerbehafteter PCR-Verfahren zu erreichen.

Bei der linearen Kombination von Formencodes können heterologe wie auch zumindest teilweise homologe Formencodes verwendet werden, die genotypisch durch einen Formencode natürlichen oder künstlichen Ursprungs definiert werden.

Unter natürlichem Ursprung wird erfindungsgemäß verstanden, daß auf schon vorhandene genetische Informationen zurückgegriffen wird, wie sie beispielsweise im Genom von Organismen niedergelegt ist. Unter Formcodes künstlichen Ursprungs wird auf Nucleinsäureebene erfindungsgemäß insbesondere verstanden, daß Sequenzen durch Algorithmen mittels Datenverarbeitungsanlagen generiert werden können, um anschließend nach diesen Anweisungen chemisch synthetisiert zu werden. Schließlich ist auch eine Herstellung durch de novo-Synthese möglich, indem Polymerasen mit den dazugehörigen Substraten wie Nucleotiden umgesetzt werden. Dabei kann die Polymerasereaktion matrizenabhängig oder -unabhängig durchgeführt werden.

Die Formenelemente und Funktionselemente, wie sie gemäß der Erfindung insbesondere verstanden werden, sind beispielsweise in Proteinen oder Peptiden als Phänotypen auffaßbar. Die entsprechenden Genotypen, beispielsweise auf Nucleinsäureebene, sind dazu die entsprechenden Form- und Funktionscodes. Bleibt man zum Beispiel auf der Nucleinsäureebene so wird der "Phänotyp" mit Funktionselementen und/oder Formenelementen, zum Beispiel durch ein Ribozym verkörpert, welches genotypisch in einer Nucleinsäuresequenz als Formencode und/oder Funktionscode entsprechend reflektiert wird. Dies bedeutet, daß erfindungsgemäß die Begriffe Funktions-/Formenelement (Phänotyp) stets quasi als komplementär mit dem Begriff Funktions-/Formencode (Genotyp) verstanden werden.

Die Formenelemente und/oder Funktionselemente bzw. Formencodes und/oder Funktionscodes können, sofern sie Nucleinsäuren sind, durch verschiedene Verfahren gewonnen werden, wie dies weiter oben angegebenen ist, nämlich durch Rückgriff auf schon bekannte Nucleinsäuresequenzen, durch Generierung künstlicher Sequenzen in Datenverarbeitungsanlagen oder durch de novo-Synthese.

Die Figur 8 erläutert die Begriff Sequenzraum, Formenraum und Funktionsraum. Analog der betrachteten Beziehung von Formenraum und Sequenzraum gilt für die Beziehung von Formenraum und Funktionsraum, daß eng benachbarte, homologe Elemente im Formenraum im Funktionsraum weit voneinander entfernt sein können. Wie in Figur 8 schematisch angedeutet, ist für die Funktion eines Polymers die Geometrie und die physicochemische Topologie und Dynamik der Moleküloberfläche maßgebend, die mit einem zweiten Molekül in Wechselwirkung tritt. Die darunterliegende Struktur, definiert aus dem Formencode, könnte sehr unterschiedlicher chemischer Natur sein. Ähnliche Funktionen im Funktionenraum erklären sich durch ähnliche Grenzflächen-Topologien.

Gerade im Hinblick auf die in Experimenten realisierbaren relativ kleinen Molekülpopulationen, ist es von entscheidender Bedeutung, daß die erzeugte Variation im Formenraum in viel effizienterer Weise als etwa die Variation im Sequenzraum die mögliche Funktionenvielfalt im Funktionsraum repräsentiert.

Die folgenden Figurenbeschreibungen erläutern an Beispielen schematisch die Erfindung näher.

Die Figur 1 betrifft zwei einzelsträngige DNA bzw. RNA-Moleküle, die chemisch oder enzymatisch (z.B. T4 RNA Ligase) ligiert werden, wobei eines der Moleküle über einen spaltbaren Linker (z.B. Biotin-Streptavidin) an fester Phase immobilisiert ist, während das andere Molekül frei in Lösung vorliegt.

Es stehen dazu heute eine ganze Reihe von Festphasenmaterialien (z.B. magnetische, oberflächenaktivierte Kunststoffkugeln) zur Verfügung. Dieses Verfahren gestattet den schrittweisen Aufbau von größeren DNAs bzw. RNAs. Nach jedem Ligationsschritt werden nicht umgesetzte RNAs weggeschüttet und die an fester Phase befindlichen Ligationsprodukte in

den nächsten Ligationsansatz transferiert. Vorteilhafterweise ist die Handhabung, insbesondere die Reinigung der jeweiligen Ligationsprodukte sehr einfach.

Nach Abschluß der letzten Ligation wird das Produkt direkt als Effektor-Molekül eingesetzt oder in einer (in vitro) Translationsreaktion zunächst in die entsprechende Proteinstruktur übersetzt, welche dann als Effektor-Molekül fungiert.

Die Figur 2 betrifft zwei vollständig doppelsträngige DNA-Moleküle, die chemisch oder enzymatisch (z.B. T4 DNA Ligase) "blunt end" ligiert werden, wobei eines über einen spaltbaren Linker an fester Phase immobilisiert ist, während das andere frei in Lösung vorliegt. Auf diese Weise können schrittweise größere doppelsträngige DNA Moleküle aufgebaut werden. Die gerichtete Ligation wird durch unterschiedliche Phosphorylierung der Reaktionspartner erreicht. Modul A und das letzte Modul sind so entworfen, daß sie jeweils eine Schnittstelle für ein Restriktionsenzym enthalten. Dies ermöglicht erstens die Abspaltung des Produktes von der festen Phase und zweitens die anschließende, gerichtete Klonierung der DNA (siehe auch Figur 5).

Zu Figur 3 : DNA-Moleküle können gemäß Figur 2 ebenfalls ligiert werden, wenn das in Lösung befindliche Molekül an einer Seite ein einzelsträngiges Ende besitzt, d.h. nicht vollständig doppelsträngig vorliegt. Dieses Ende steht auf diese Weise nicht für die Doppelstrang-spezifische Ligation, z.B. mit T4 DNA Ligase zur Verfügung. In Kombination mit den schon erwähnten Phosphorylierungsstrategien (Figur 2, insbesondere Variante 1) ergibt sich die Möglichkeit, die Ligation ohne unerwünschte Nebenprodukte durchzuführen. Das in Lösung befindliche DNA-Molekül kann so entworfen sein, daß es vor seinem einzelsträngigen Ende noch die Schnittstelle eines Restriktionsenzymes vorzugsweise die eines Class IIS Enzyms (z.B. AlwI) mit Erkennungsstelle in dem abzuschneidenden, teilweise einzelsträngigen DNA-Stück) besitzt. Nach der

Ligation kann das Ligationsprodukt an fester Phase mit dem Restriktionsenzym geschnitten werden. Auf diese Weise entsteht wieder ein vollständig doppelsträngiges DNA-Molekül an fester Phase. Alternativ kann das einzelsträngige Ende mit einer Polymerase zum Doppelstrang aufgefüllt oder mit einer Exonuclease abverdaut werden.

Zu Figur 4: Restriktionsschnittstellen können (überlappend) auch entstehen, indem zwei doppelsträngige DNA-Moleküle miteinander ligiert werden.

Zu Figur 5: Vollständig oder teilweise doppelsträngige DNA-Moleküle können gemäß Figuren 1 - 4 ligiert werden, auch wenn Mischungen von Molekülen (z. B. B, C, D) verwendet werden. Auf diese Weise entstehen Mischungen von immobilisierten Molekülen, die jeweils verschiedenen Kombinationen der eingesetzten Bausteine entsprechen. Am Ende des letzten Ligationsschrittes kann die Gesamt-DNA oder ein Teil davon mit Hilfe von Restriktionsenzymen, die innerhalb des Konstruktions schneiden, von der festen Phase abgespalten und ggf. in ein Phagen- oder Bakteriendisplay-System kloniert werden. Die DNA kann aber auch in einem kombinierten in vitro Transkriptions- und Translationssystem exprimiert werden.

Zu Figur 6: Ausgehend von Modul-Bibliotheken können Peptide, Proteindomänen und kleine Proteine durch zufällige Kombination von einzelnen Modulen erzeugt werden. Entsprechend einem hierarchischen Verfahren zum Proteindesign können in einer weiteren Stufe dann auch Proteindomänen als Bausteine kombiniert werden. Auf jeder Komplexitätsstufe können Mutationen eingefügt werden, die - ohne die globale Struktur zu verändern - eine Feinabstimmung der dreidimensionalen Anordnung chemischer Gruppen erlauben.

Figur 7 erläutert schematisch, daß verschiedene Proteine trotz unterschiedlicher, katalytisch aktiver Aminosäuren im aktiven Zentrum in Bezug auf das Substrat homologe Funktionen

besitzen (Chymotrypsin/Trypsin) oder trotz ähnlicher räumlicher Anordnung der Aminosäuren im aktiven Zentrum gänzlich unterschiedliche Reaktionen katalysieren (Trypsin/Elastase) können.

Figur 8 erläutert den Zusammenhang der Begriffe Sequenzraum-Formenraum-Funktionsraum.

Fig. 9 - 11

Oligomere oder Polymere werden durch matrizenabhängige, enzymatische oder chemische Synthese durch Verlängerung von stochastischen (randomisierten) oder ausgewählten (konstruierten) Primer-Molekülen hergestellt. Die Primer können komplementär zu 1.) diskreten Sequenzen ausschließlich am Ende des ursprünglichen Matrizen-Moleküls (Figur 9/10) diskreten Sequenzen überall im ursprünglichen Matrizen-Molekül sein (Figur 10/11) aus einer Mischung von Zufallssequenzen bestehen, die die Synthese je nach (teilweiser) Komplementarität zufällig an vielen Stellen beginnen lassen (Fig. 10).

Entweder die Primer oder, wie in der Figur gezeigt, auch die Matrizen-DNA können zur Vereinfachung späterer Aufreinigungsprozeduren biotinyliert sein. Dies würde insbesondere bei der Strangtrennung (z. B. an Streptavidin-Dynabeads) zur Aufreinigung der verlängerten Primer günstig sein.

Fig. 12

Statt normaler dNTPs (Desoxy-Nucleosidtriphosphate) werden insbesondere thio-NTPs eingesetzt. Kettenabbruchmoleküle können dann zum Beispiel normale ddNTPs (Didesoxy-Nucleosidtriphosphate) sein. Es ist bekannt, daß Phosphodiesterbindungen leicht durch Exonuclease III in 50 mM Tris/HCl, 5 mM MgCl<sub>2</sub>, bei pH 10,0 spezifisch in Minuten gespalten werden können. Thiophosphat-Bindungen werden dagegen nicht gespalten (Labeit et al., DNA 5:173, 1986). Auf diese Weise kann man nach Inaktivierung der Polymerase die Enden der

- 17 -

entstandenen Polymere durch enzymatisches Entfernen der ddNTPs "entschützen".

Fig. 13

Die entstandenen und entschützten Polymere können entweder thermisch oder chemisch, z. B. mit NaOH, getrennt werden, wobei sich die Biotin-gekoppelten Moleküle z. B. an Streptavidin-Dynabeads abtrennen lassen. Nach Einstellen physiologischer Bedingungen hybridisieren die entschützten Polymere zu teilweise überlappenden Duplices.

Fig. 14

PCR ohne Primer führt zur eigentlichen (Re-)Kombination der Polymere und zur weiteren Verlängerung derselben. Nachdem die teilweise überlappenden Duplices vervollständigt wurden, findet eine weitere PCR mit (endständigen) Primern statt, die wieder Produkte der ursprünglichen Länge erzeugt. Darunter befinden sich auch solche Sequenzen, in denen mehrere Marker vereinigt, neu kombiniert sind.

Der Sequenzraum ist durch die linearen Nachbarschaftsbeziehungen der Polymer-Bauelemente einer Polymerstruktur definiert. Homologien beschreiben Ähnlichkeiten (in %) in der Abfolge der Bauelemente einer chemischen Stoffklasse. Je höher der Verwandtschaftsgrad zweier Sequenzen desto geringer der Abstand im Sequenzraum.

- a) ... AATAATGCGCAATATTAGGCCT...
- b) ... AATAAAAAGCAATATTAAGCCT...
- c) ... TTAGCTAGCGATGCGCGCCGGG ...

Zum Beispiel weisen die Sequenzen a) und b) eine erhebliche Homologie auf, während Sequenz c) keinerlei Ähnlichkeiten mit a) und b) zeigt.

Der Formenraum ist definiert durch die "räumlichen" Nachbarschaftsbeziehungen der durch ihn repräsentierten Polymere.

Der Abstand zweier Sequenzen ist durch den Verwandtschaftsgrad ihrer Strukturen bestimmt. Homologie bedeutet hier Ähnlichkeit der Gesamtstrukturen von Polymeren, die wiederum aus chemisch verknüpften Bauelementen bestehen. Im Formenraum benachbarte Moleküle können im Sequenzraum durchaus weit voneinander entfernt liegen und umgekehrt. [Analog s. o.: Struktur a) 3 alpha-Helices, Struktur b) 2 Alpha-Helices plus unstrukturierter Bereich mit endständiger, kurzer Helix, c) antiparalleles beta-Faltblatt aus 4 Blättern] Der Funktionenraum ist definiert durch die geometrische, dynamische und physikalisch/chemische Oberflächenstruktur, die mit einem weiteren Molekül in spezifische Wechselwirkung treten kann. Homologien beschreiben Ähnlichkeiten der Oberflächenstruktur und den damit verbundenen Wechselwirkungseigenschaften.

Im folgenden werden lineare Kombinationen von Formencodes über in vitro Rekombination von Formencodes aus natürlichen oder in vitro hergestellten Mutanten beschrieben.

Bei der linearen Kombination von Formencodes können heterologe wie auch zumindest teilweise homologe Formencodes verwendet werden. Sequenzhomologe Formencodes können zufällig in einer zu rekombinierenden Mischung vorhanden sein oder bewußt ausgewählt werden. Diese Mischung kann beispielsweise, wie bei Eigen & Henco WO 92/18645 beschrieben, homolog auseinander hervorgegangene Mutantenkollektive einer Ausgangssequenz, oder homologe Gene verwandter oder unterschiedlicher Organismen enthalten. Dabei können ähnliche Sequenzen, z. B. bezüglich ihrer Funktionscodes, sehr unterschiedlich sein.

Die Natur hat für ein und dieselbe oder sehr ähnliche Reaktionen, für verschiedene Wirtssysteme ähnliche oder molekular unterschiedliche Enzyme evolviert, von denen angenommen werden kann, daß sie für die jeweilige Umgebung, für die sie angepaßt wurden, optimale Lösungen bieten. Dafür ist die Penicillinacetylase beispielhaft. Dieses Enzym ist für die

industrielle Anwendung im Bereich der Synthese von zentraler Bedeutung. Vor einer Synthese halbsynthetischer Derivate des Penicillin-Grundkörpers muß in natürlicher Weise synthetisiertes Penicillin zunächst schonend einer Acylase-Reaktion unterworfen werden, bevor es in einer Umkehrung des Prozesses wiederum mit künstlichen Derivaten reacyliert werden kann. Für beide Reaktionen kann die Penicillinacylase eingesetzt werden. Für diese Reaktion sind bestimmte Reaktionsbedingungen und Substrate erwünscht. Diese Bedingungen unterscheiden sich jedoch von der *in vivo* Situation des Mikroorganismus, aus dem jeweils Penicillinacylasen isoliert wurden. Dies trifft zum Beispiel auf die optimale Lage des Gleichgewichtes für das acylierte Syntheseprodukt oder für die hydrolytische Spaltung zu. Es ist gewünscht, das Enzym bezogen auf die Umsatzzahl unter den industriell am besten geeigneten Bedingungen zu optimieren.

Insbesondere kann von einem Gen einer bestimmten natürlich vorkommenden Acylase ausgegangen und diese konsekutiven Mutations- und Selektionszyklen unterworfen werden. Wenn verschiedene, aktive Mutanten gefunden sind, lassen sich die als positiv selektierten unterschiedlichen Varianten bezogen auf die jeweils selektierten Punktmutationen über Rekombination ein weiteres Mal durchmischen. Die so zu mischenden Varianten können wie in WO 92/18645 beschrieben gewonnen werden. Die Natur verfügt häufig bereits über eine Kollektion von positiv selektierten Mutanten in Form der Enzymgene aus verschiedenen Mikroorganismen, die beispielsweise den gewünschten Reaktionstyp katalysieren. Von diesen Kollektionen ausgehend, lassen sich bereits Spektren rekombinierter Formencodes und Funktionscodes erzeugen, bevor eventuell wieder im weiteren Verlauf Mutations- oder eine Kombination von Mutations-/Rekombinationszyklen durchlaufen werden. Es ist durchaus vorteilhaft, am Anfang eines solchen Prozesses von möglichst umfangreichen Formcodes auszugehen, deren Mutationen sich in einem bestimmten Kontext des jeweiligen Gens als positiv oder neutral erwiesen haben.

Die *in vitro*-Rekombination wird dabei vorzugsweise nach zwei unterschiedlichen Strategien durchgeführt.

Rekombination kann ein in der Regel unerwünschtes Nebenprodukt einer Amplifikationsreaktion im Sinne einer PCR-Reaktion sein. Wenn in der Sättigungsphase einer PCR-Reaktion nach dem Durchlaufen vieler Zyklen die Lösung an Reagentien oder Enzym verarmt und die Reaktion unterhalb des  $K_m$ -Wertes für bestimmte Nucleotidtriphosphate verläuft, kommt es zwangsläufig zu nicht fertiggestellten Syntheseprodukten. Solche Ereignisse sind als unerwünschte Artefakte bereits von Simon Wain-Hobson diskutiert worden, um HIV-Varianten als mögliche Artefakte nach erfolgter PCR-Amplifikation zu beschreiben. Dieser Effekt wird aber erfindungsgemäß eingesetzt und gesteuert, insbesondere noch verstärkt, daß wunschgemäß unvollständig synthetisierte Produkte dominant werden. Wenn gleichzeitig die Primer-induzierte Neusynthese unvollständig erfolgt, hybridisieren maßgeblich unvollständige Syntheseprodukte mit vollständig oder ebenfalls unvollständig synthetisierten Gegensträngen. Dabei kommt es zu molekularen Rekombinationsereignissen, bei denen verschiedene Gensegmente im Sinne einer Rekombination von Formencodes miteinander rekombiniert werden.

Erfindungsgemäß lässt sich nach einer weiteren spezifischen Vorgehensweise die Rekombination während und nicht nur nach einer PCR-Reaktion steuern. Hierbei werden der standardmäßigen PCR-Reaktion kurze Oligomere zugesetzt, die nur dann als PCR-Primer fungieren, wenn die Initiation der Synthese mittels thermostabiler Polymerase bei vergleichsweise tiefen Temperaturen erfolgt. Wann immer die korrekte PCR-Reaktion dominieren soll, wird ein normaler Temperaturzyklus ausgeführt. Wenn es zu internen Startreaktionen kommen soll, werden einige Zyklen bei niedriger Temperatur initiiert, eventuell unter Zusatz von Polymerasen wie DNA-Polymerase I, wie sei bei Oligomer-gestarteten Markierungsreaktionen eingesetzt wird (Sambrook, Fritsch, Maniatis, "Molecular

cloning"). Die entstehenden unvollständigen Sequenzen können sich bei weiteren Amplifikationsrunden zusammenlagern, wobei die überhängenden Enden jeweils am 3'-Ende aufgefüllt werden können. Bei diesen Reaktionen muß gemäß an sich bekannten Reassoziationskinetik von Nucleinsäuren darauf geachtet werden, daß die für eine Rekombination bestimmten Sequenzen in hinreichender Konzentration zur Verfügung stehen, um in wenigen Sekunden bis Minuten die unvollständig gepaarten Duplices auszubilden. Um zu vermeiden, daß bei der matrizenvermittelten Neusynthese unerwünscht eine Strangverdrängung anstelle eines Rekombinationseignisses stattfindet, werden insbesondere solche Polymerasen verwendet, die keine Strangverdrängung induzieren oder keine 5'-3'-Exonucleaseaktivität aufweisen. Stattdessen lassen sich bevorzugt thermostabile Ligasen einsetzen, so daß Rekombinationseignisse durch kovalente Verknüpfung der Fragmente fixiert werden.

Im erfindungsgemäßen Verfahren zur Rekombination von Formencodes setzt man Elemente mit zumindest teilweisen Sequenzhomologien, wie sie oben beschrieben wurden, ein. Mit Hilfe matrizenabhängiger chemischer oder enzymatischer DNA- oder RNA-Synthese durch Verlängerung von erzeugten (randomisierten) Primern oder ausgewählten (konstruierten) Primern wird eine Vielzahl von Fragmenten mindestens einer ursprünglichen Sequenz erzeugt (s. Fig. 9 - 11). Ausgewählte Primer mit definierter Sequenz können dabei so positioniert werden, daß bestimmte Bereiche der zu bearbeitenden DNA- oder rRNA-Moleküle, z. B. aktiven Zentren, Endonuclease-spezifische Spaltstellen oder genregulatorische Elemente, vom Rekombinationsprozeß ausgeschlossen sind und somit unverändert erhalten bleiben. Der Einsatz teilweise randomisierter Primer in Bereichen (partieller) Komplementarität analog Mutagansierungs-Primern, kann dazu verwendet werden, zusätzlich eine erhöhte Mutationsrate einzuführen. Durch den Einsatz einer kleinen, subinhibitorischen Menge von Kettenabbruchmonomeren, in der DNA-Synthese von vorzugsweise Dideoxynucleotiden, wird ein zufälliger Abbruch der Verlängerungsreaktion und damit

eine Längenvarianz der synthetisierten Polymere erreicht. Mittels des Verhältnisses der Konzentration des Abbruchreagenz zu den Konzentrationen der Nucleotidmonomeren läßt sich wie bei einer Sequenzierungsreaktion die durchschnittliche Kettenlänge des Syntheseproduktes steuern. Nach Abtrennung des polymerisierenden Agenz, beispielsweise einer Inaktivierung des Enzyms, kann die endständige Schutzgruppe, d. h. das Kettenabbruchmonomer ganz oder teilweise wieder abgespalten werden, damit die entstehenden Polymere wieder gute Substrate für die Verlängerungsreaktion sind (Fig. 12). Die so entschützten DNA- oder RNA-Polymere werden dann mindestens einem Zyklus aus Denaturierung/Hybridisierung teilweise komplementärer Stränge gefolgt von einer Auffüllreaktion unterzogen. Am Schluß des Verfahrens wird das entstandene Gemisch verlängerter Polymere einer Polymerase-Kettenreaktion unterzogen, wobei die Primer vorzugsweise komplementär zu den Enden der ursprünglich eingesetzten Sequenz liegen sollten. Auf diese Weise entstehen wieder Produkte der ursprünglichen Länge. Diese enthalten jetzt aber Kombinationen von Sequenzabschnitten verschiedener vorteilhafter, bereits selektierter, einzelner Punktmutationen in sehr effizienter Weise zu kombinieren, statt sie sequentiell in einem stochastischen Prozeß erst erzeugen zu müssen

A n s p r ü c h e

1. Verfahren zur Herstellung oligomerer oder polymerer Funktionselemente aus Formenelementen, wobei die Funktionselemente erhältlich sind durch Verknüpfung von mindestens zwei Formenelementen, von denen mindestens ein Formenelement selbst aus mindestens zwei Monomeren aufgebaut ist, die durch mindestens eine chemische Bindung verknüpft sind, die der chemischen Bindung zwischen zwei Formenelementen entspricht.
2. Verfahren gemäß Anspruch 1, wobei die Verknüpfung der Formenelemente unter Einsatz einer festen Phase als Reaktionsträger durchgeführt wird.
3. Verfahren gemäß Anspruch 1 und/oder 2, wobei die Verknüpfung der Formenelemente chemisch und/oder enzymatisch erfolgt.
4. Verfahren gemäß mindestens einem der Ansprüche 1 - 3, wobei die Verknüpfung von Formenelementen zu Funktionselementen planmäßig und/oder stochastisch erfolgt.
5. Verfahren gemäß mindestens einem der Ansprüche 1 - 4, wobei die Verknüpfung schrittweise aufbauend, stereospezifisch und/oder gerichtet erfolgt.
6. Verfahren gemäß mindestens einem der Ansprüche 1 bis 5, dadurch gekennzeichnet, daß die Formenelemente zur Stoffklasse der Nucleinsäuren, doppelsträngiger und/oder einzelsträngiger DNA und/oder RNA und/oder modifizierten Nucleinsäuren und/oder Peptiden und/oder Polypeptiden gehören und/oder aus sonstigen kopplungsfähigen chemischen Oligomer-Formenelementen aufgebaut sind.
7. Verfahren gemäß mindestens einem der Ansprüche 1 - 6, dadurch gekennzeichnet, daß die Formenelemente als bereits synthetisierte Oligomer-Bausteine eingesetzt werden oder zunächst im Reaktionsgefäß generiert werden.

8. Verfahren gemäß mindestens einem der Ansprüche 1 - 7, dadurch gekennzeichnet, daß die Reaktionen in parallel geführten Micro-Reaktionsansätzen durchgeführt werden, bei denen Formenelemente in vorbestimmter Reihenfolge verknüpft werden.
9. Verfahren gemäß mindestens einem der Ansprüche 1 - 8, dadurch gekennzeichnet, daß nach erfolgter Synthese die Reaktionsprodukte, wie Funktionselemente oder Vorstufen davon festphasengekoppelt bleiben oder in die lösliche Phase entkoppelt werden.
10. Verfahren nach Anspruch 9, wobei die Reaktionsprodukte mit einem biologischen Testsystem vereinigt werden, wobei die Funktion im gleichen Volumenelement wie die Synthese bewertend gemessen wird, z.B. durch den Einsatz der FCS--Analysetechnik.
11. Verfahren gemäß mindestens einem der Ansprüche 1 - 10, dadurch gekennzeichnet, daß pro Reaktionsschritt bei der schrittweisen Verknüpfung der Formenelemente jeweils ein Formenelement als Reaktionspartner an fester Phase gekoppelt ist.
12. Verfahren gemäß mindestens einem der Ansprüche 1 - 11, dadurch gekennzeichnet, daß Mischungen von Formenelementen eingesetzt werden und/oder generiert werden können.
13. Verfahren gemäß mindestens einem der Ansprüche 1 - 12, dadurch gekennzeichnet, daß im Falle des Aufbaues von Nucleinsäure-Formenelementen und/oder der Verknüpfung von Nucleinsäure-Formenelementen wenigstens ein Reaktionspartner eine Schnittstelle für ein Restriktionsenzym enthält und/oder frei von Start- und/ oder Stopcodons ist.
14. Verfahren gemäß mindestens einem der Ansprüche 1 - 13, dadurch gekennzeichnet, daß über die Einführung von Restriktionsschnittstellen, insbesondere solchen für Enzyme der

- 25 -

Klasse IIS beliebige Sequenzen gerichtet verknüpft werden können, ohne daß Sequenzerfordernisse des erwünschten Endproduktes die Wahl des Reaktionsenzyms beeinflußen.

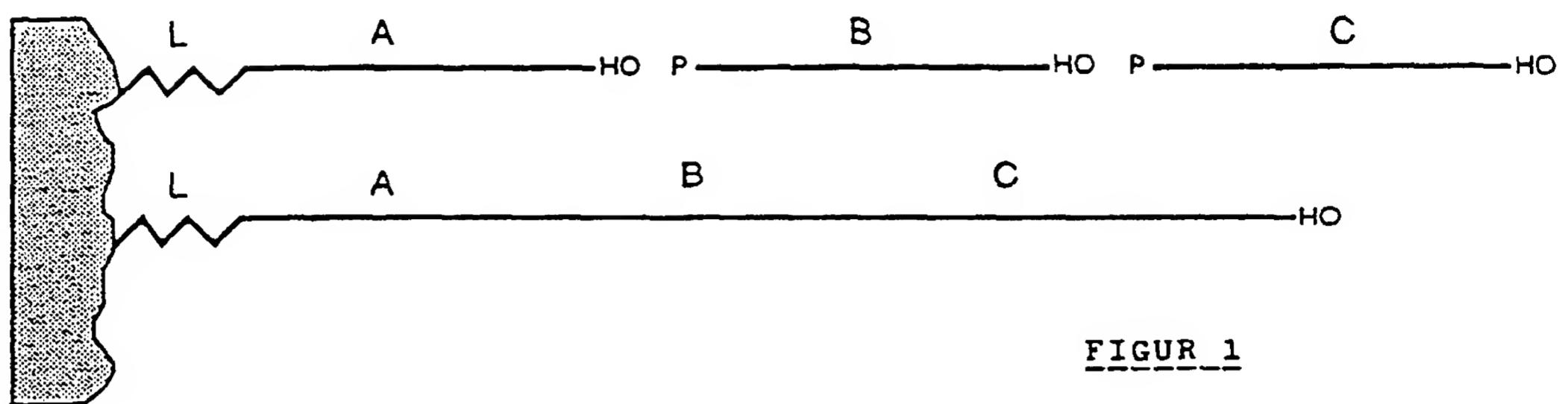
15. Verfahren gemäß mindestens einem der Ansprüche 1 - 14, dadurch gekennzeichnet, daß über die Einführung von einzelst-rängigen Überhängen und/oder selektive und reversible chemische und/oder enzymatische Modifikation der 3'-Enden und/oder der 5'-Enden der Nucleinsäuren, zum Beispiel Phosphorylierung, beliebige Sequenzen gerichtet verknüpft werden können, ohne daß dabei irgendwelche Anforderungen an die Sequenz des erwünschten Endproduktes entstehen.
16. Verfahren gemäß mindestens einem der Ansprüche 1 - 15, dadurch gekennzeichnet, daß Formenelemente nach dem Vorbild röntgenkristallografisch analysierter natürlicher Proteine oder Polypeptide eingesetzt werden.
17. Verfahren gemäß mindestens einem der Ansprüche 1 - 16, dadurch gekennzeichnet, daß mindestens eines der verwendeten Formenelemente aus Selektionsexperimenten stammt.
18. Verfahren gemäß mindestens einem der Ansprüche 1 - 17, dadurch gekennzeichnet, daß die Formenelemente zwischen 1 und 60 Aminosäuren enthalten oder Nucleotide entsprechender Kodierungslänge.
19. Verfahren gemäß mindestens einem der Ansprüche 1 - 18, dadurch gekennzeichnet, daß Formenelemente eingesetzt werden, die an bestimmten Positionen degeneriert sind und/oder Deletionen oder Insertionen tragen.
20. Verfahren gemäß einem der Ansprüche 1 bis 19, dadurch gekenn-zeichnet, daß die Funktions- und/oder Formenelemente bzw. Funktionscodes und/oder Formencodes als Oligo- oder Poly-nucleotide niedergelegt sind, die erhältlich sind

- 26 -

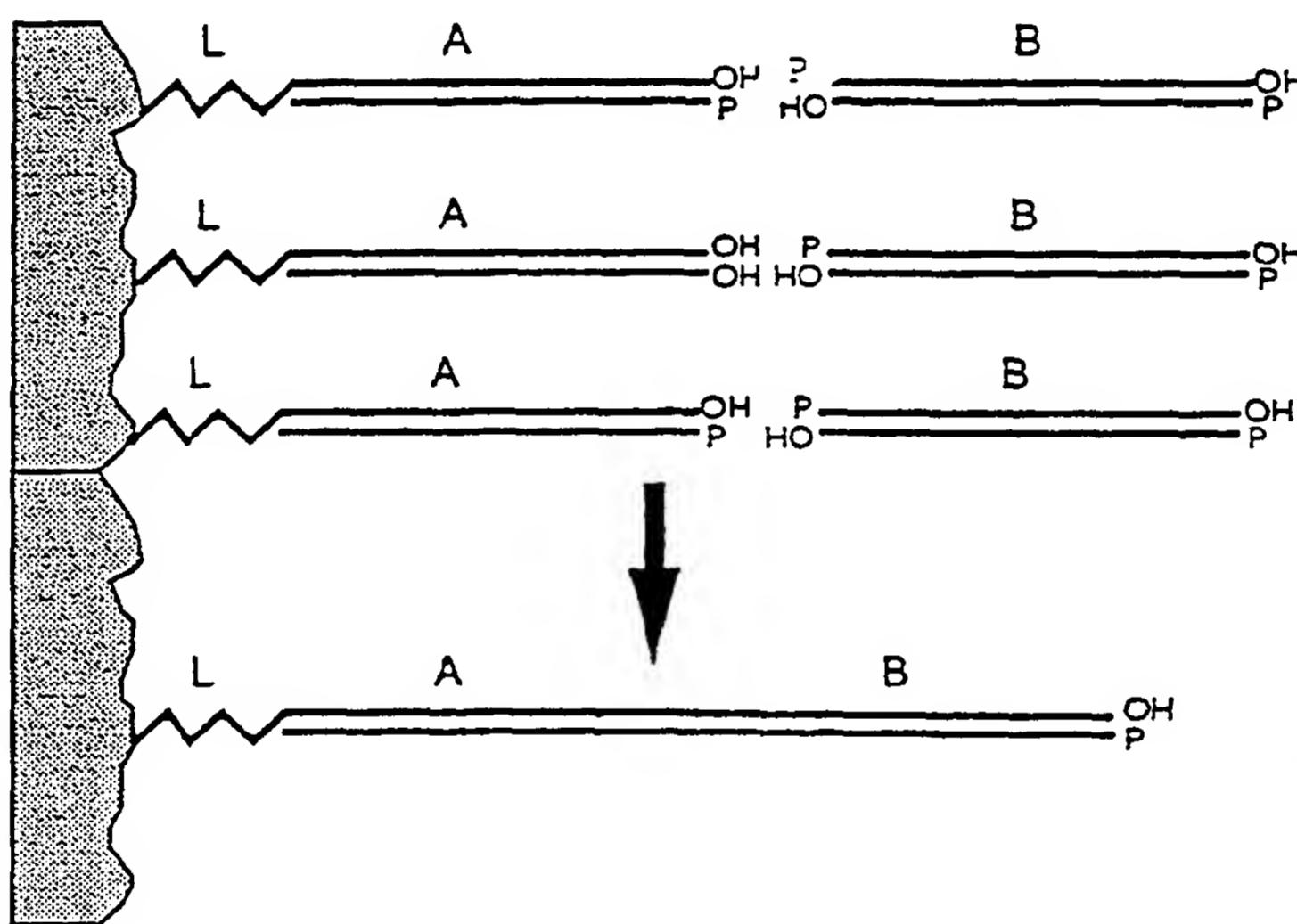
- durch Generierung aus Algorithmen, insbesondere evolutiver Algorithmen,
- durch Übernahme oder Modifizierung natürlich vorkommender Nucleinsäuren und/oder,
- Generierung mittels de novo Synthese von Oligo-/Polynucleotiden durch matrizenabhängige oder -unabhängige Reaktionen von Polymerasen mit Nucleotiden.

21. Verwendung des Verfahrens gemäß mindestens einem der Ansprüche 1 - 20 zur Synthese parallel aufgebauter Formen-Bibliotheken funktionaler Oligomere oder Polymere.

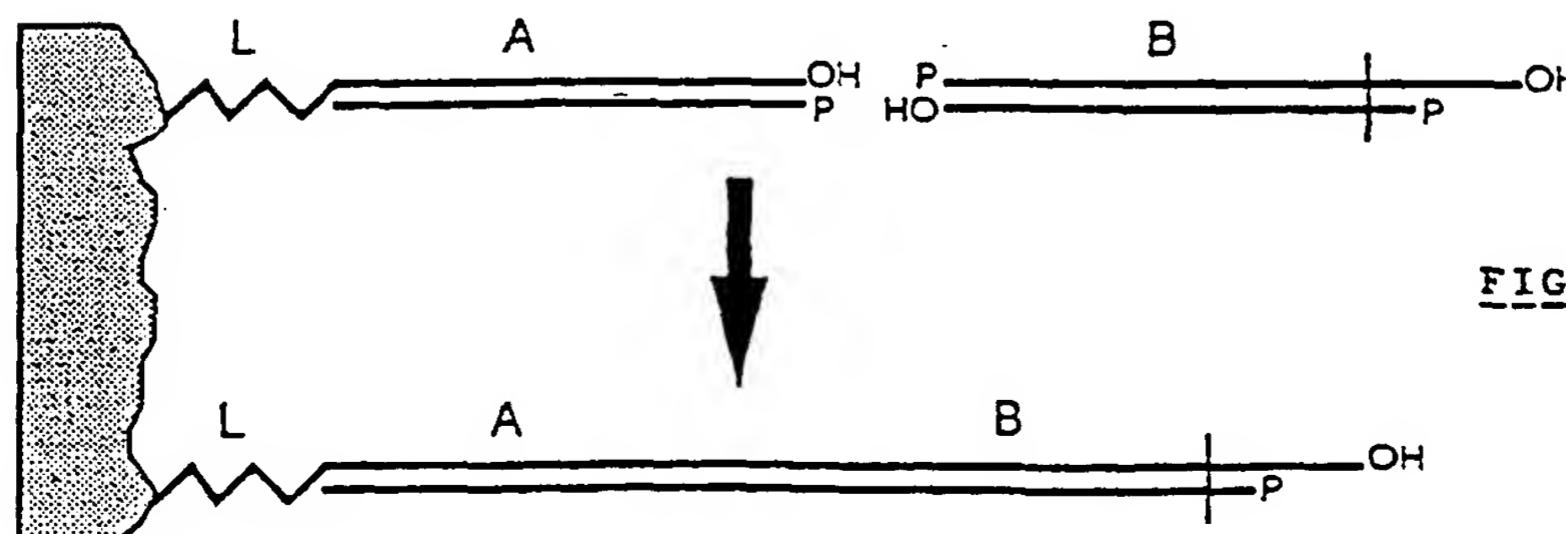
1 / 9



FIGUR 1



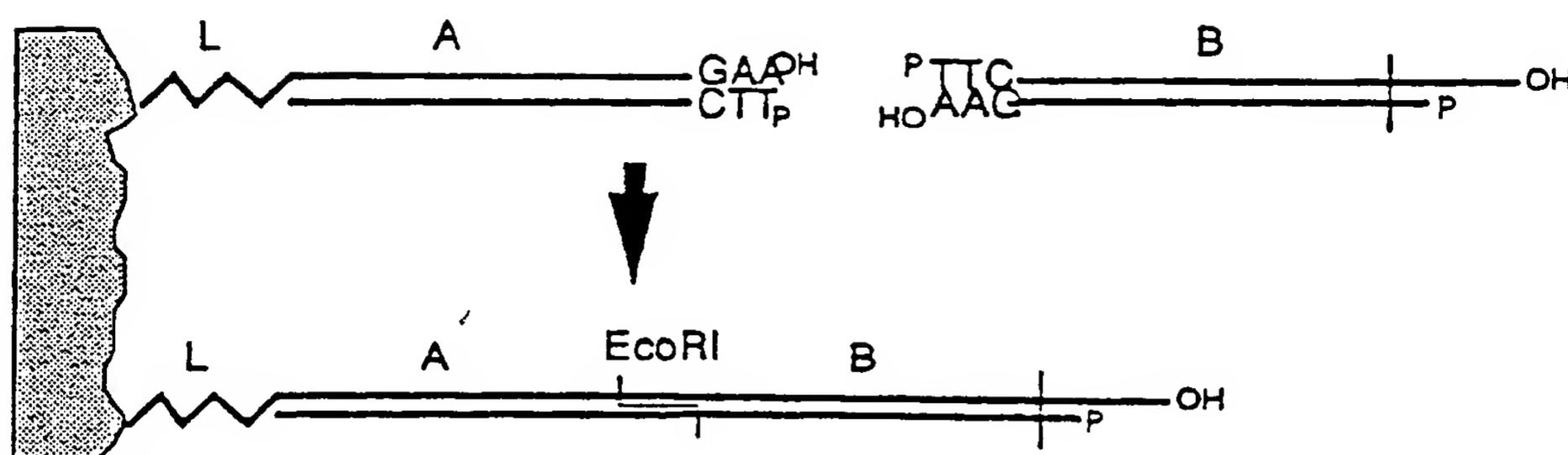
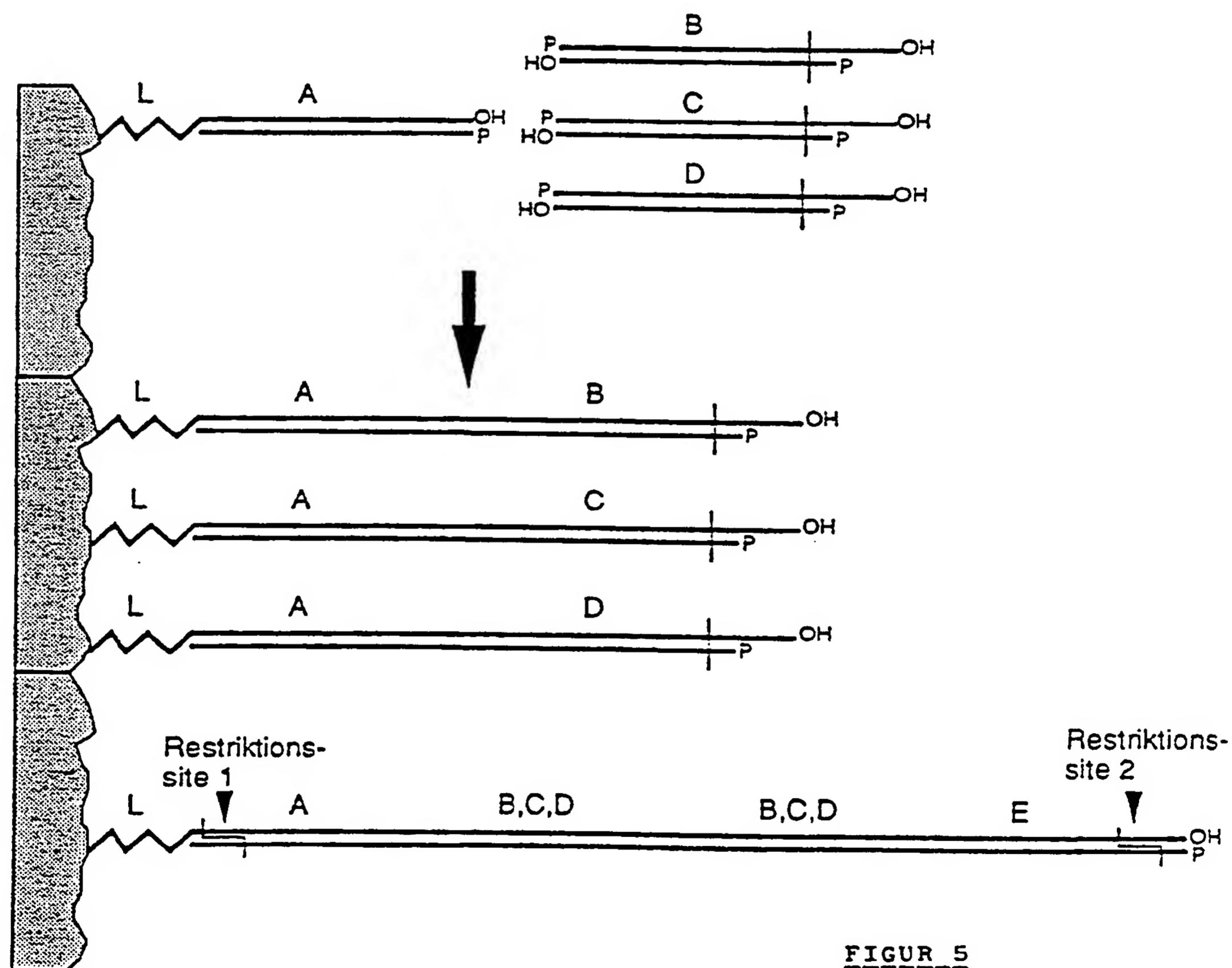
FIGUR 2



FIGUR 3

ERSATZBLATT (REGEL 26)

2 / 9

FIGUR 4FIGUR 5

ERSATZBLATT (REGEL 26)

3 / 9

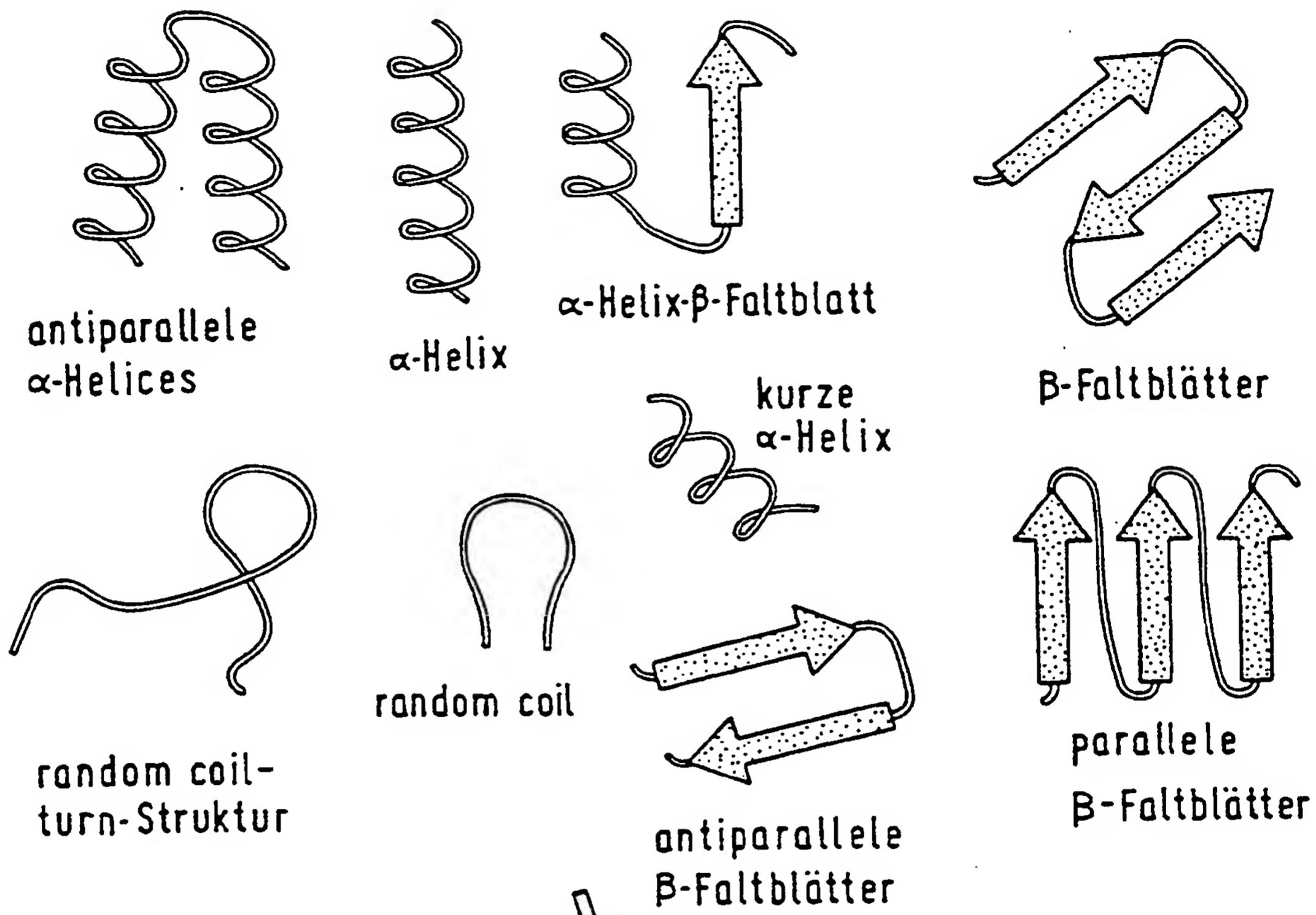
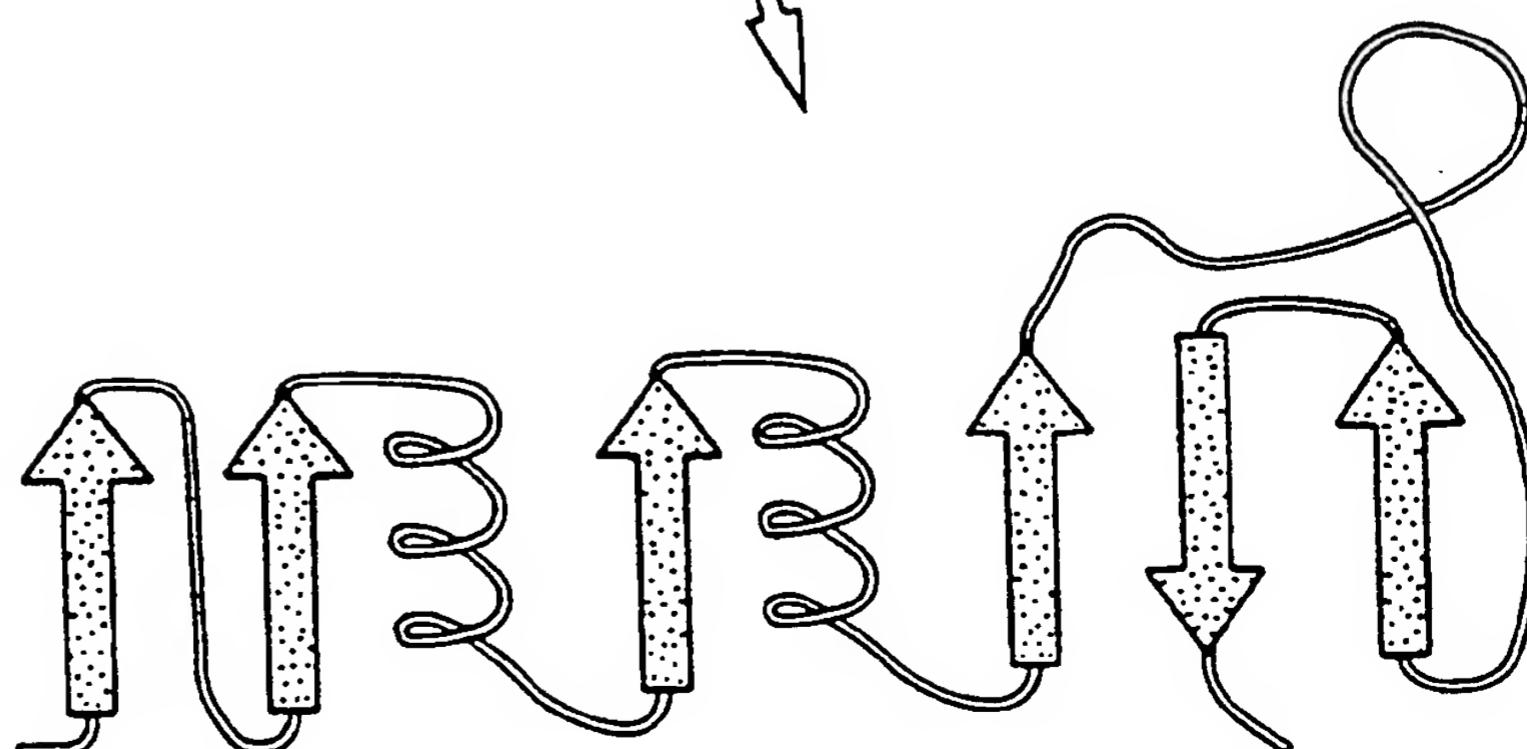
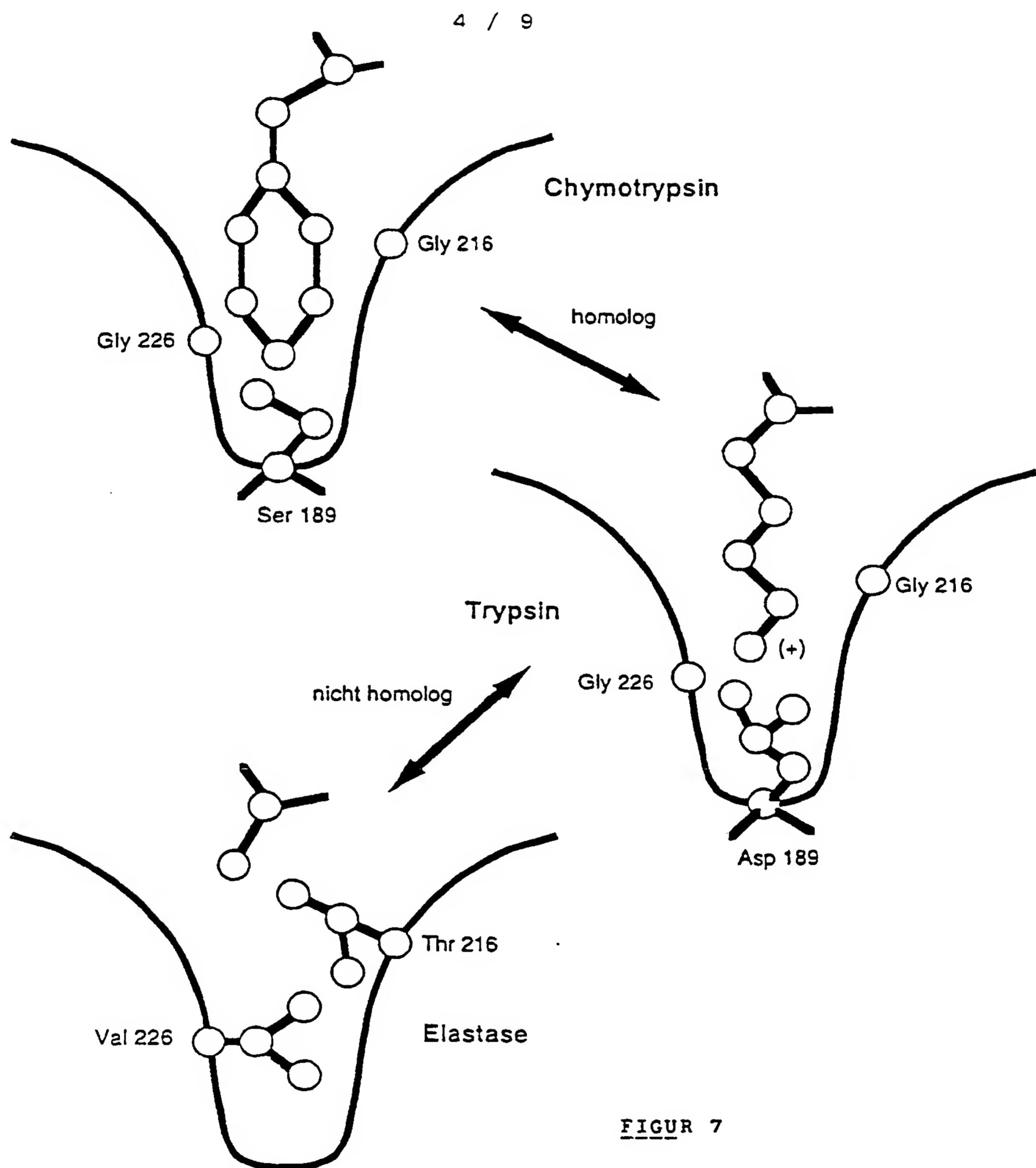


FIG. 6



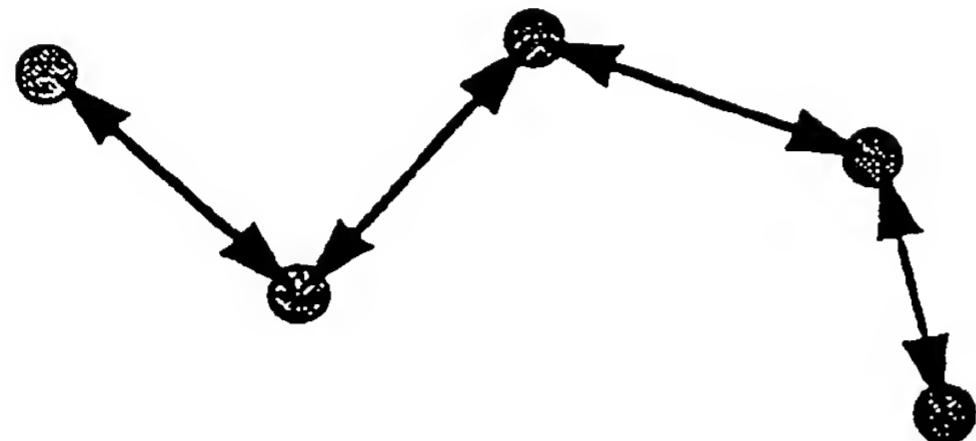
zusammengesetzte, komplexe Struktur  
ERSATZBLATT (REGEL 26)



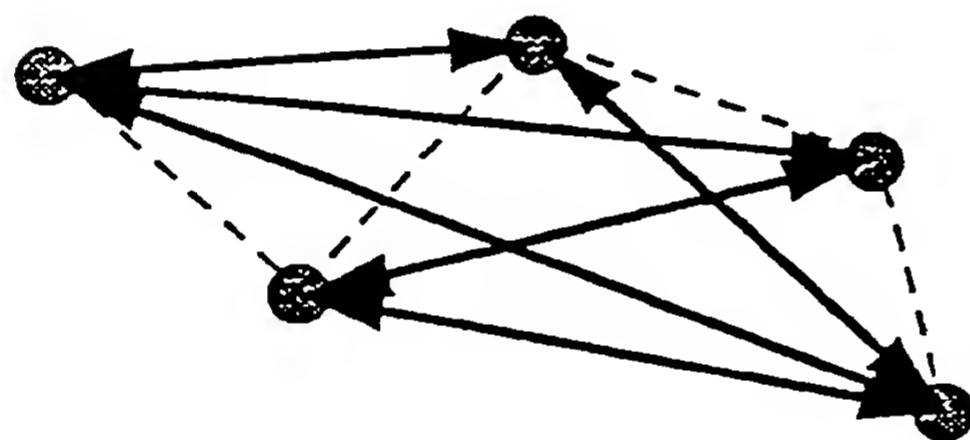
ERSATZBLATT (REGEL 26)

5 / 9

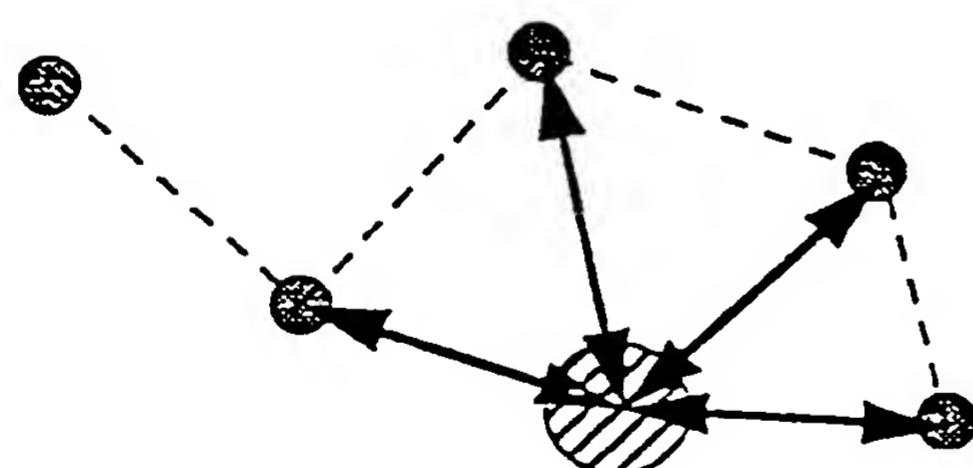
## Sequenzraum - Formenraum - Funktionsraum



Der Sequenzraum ist durch die linearen Nachbarschaftsbeziehungen einer Polymerstruktur definiert. Homologien beschreiben Ähnlichkeiten (in %) in der Abfolge der Bauelemente einer chemischen Stoffklasse.



Der Formenraum ist definiert durch die räumlichen Nachbarschaftsbeziehungen der Bausteine einer Polymerstruktur. Homologien beschreiben Ähnlichkeiten der Raumstruktur von nicht unmittelbar chemisch verknüpften Bauelementen.

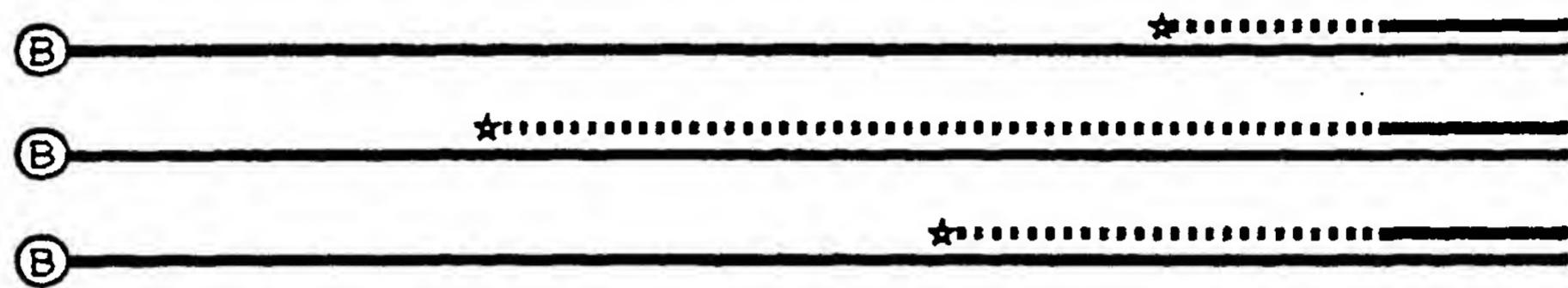
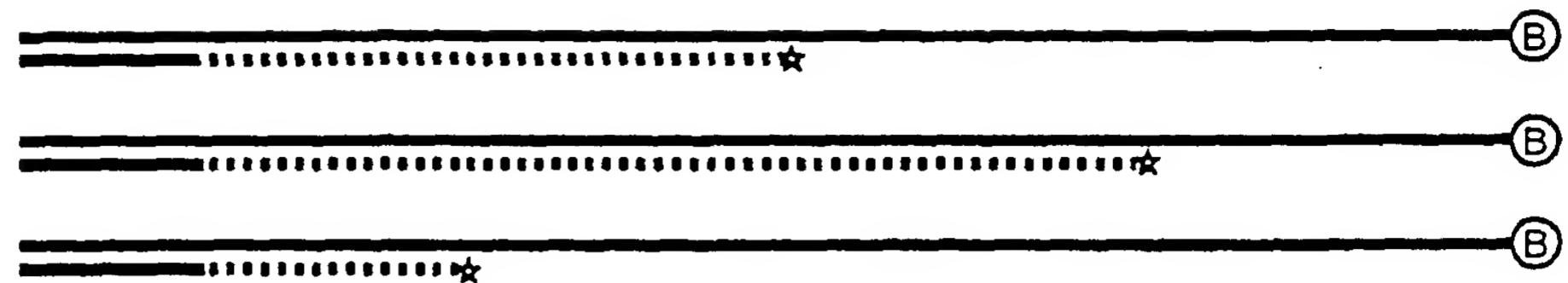


Der Funktionsraum ist definiert durch die geometrische, dynamische und physikalisch/chemische Oberflächenstruktur, die mit einem weiteren Molekül in spezifische Interaktion treten kann. Homologien beschreiben Ähnlichkeiten der Oberflächenstruktur und den damit verbundenen Wechselwirkungseigenschaften.

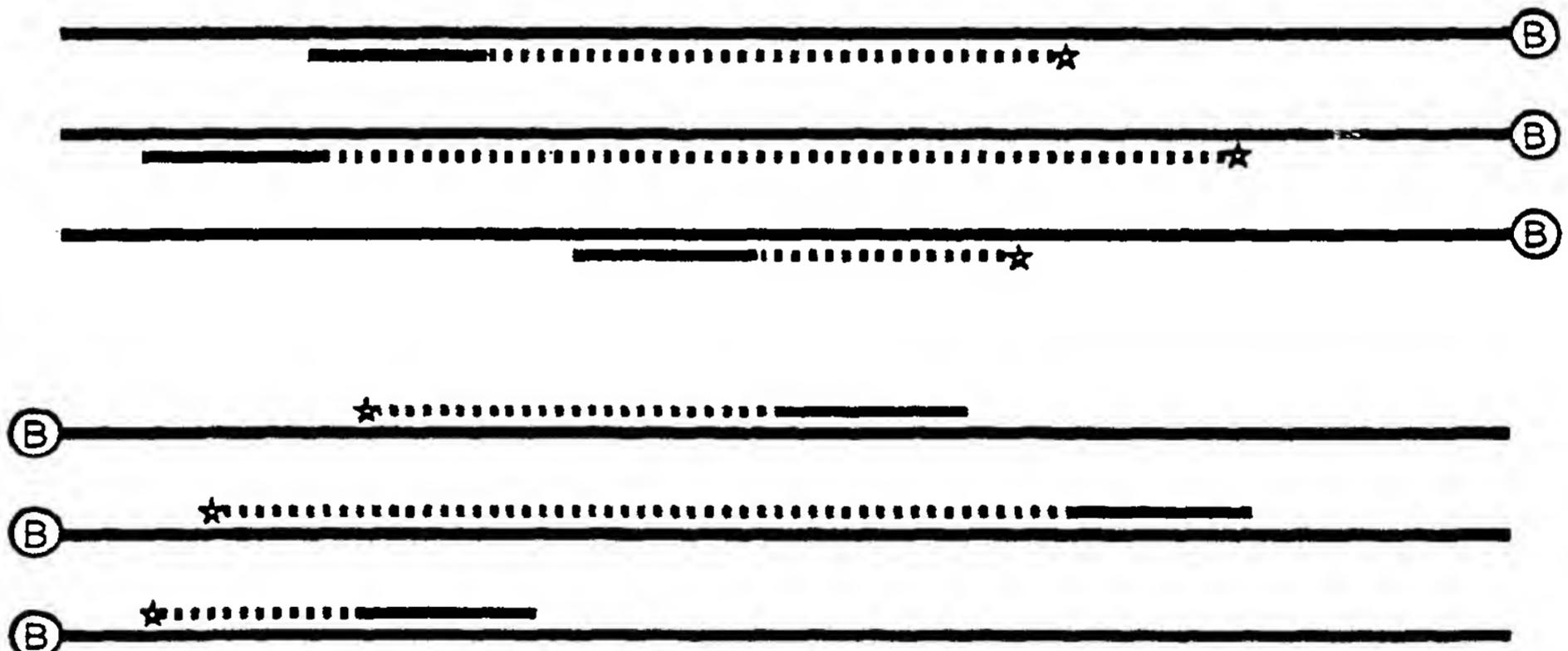
EIGUR\_8

ERSATZBLATT (REGEL 26)

6 / 9



Figur 9

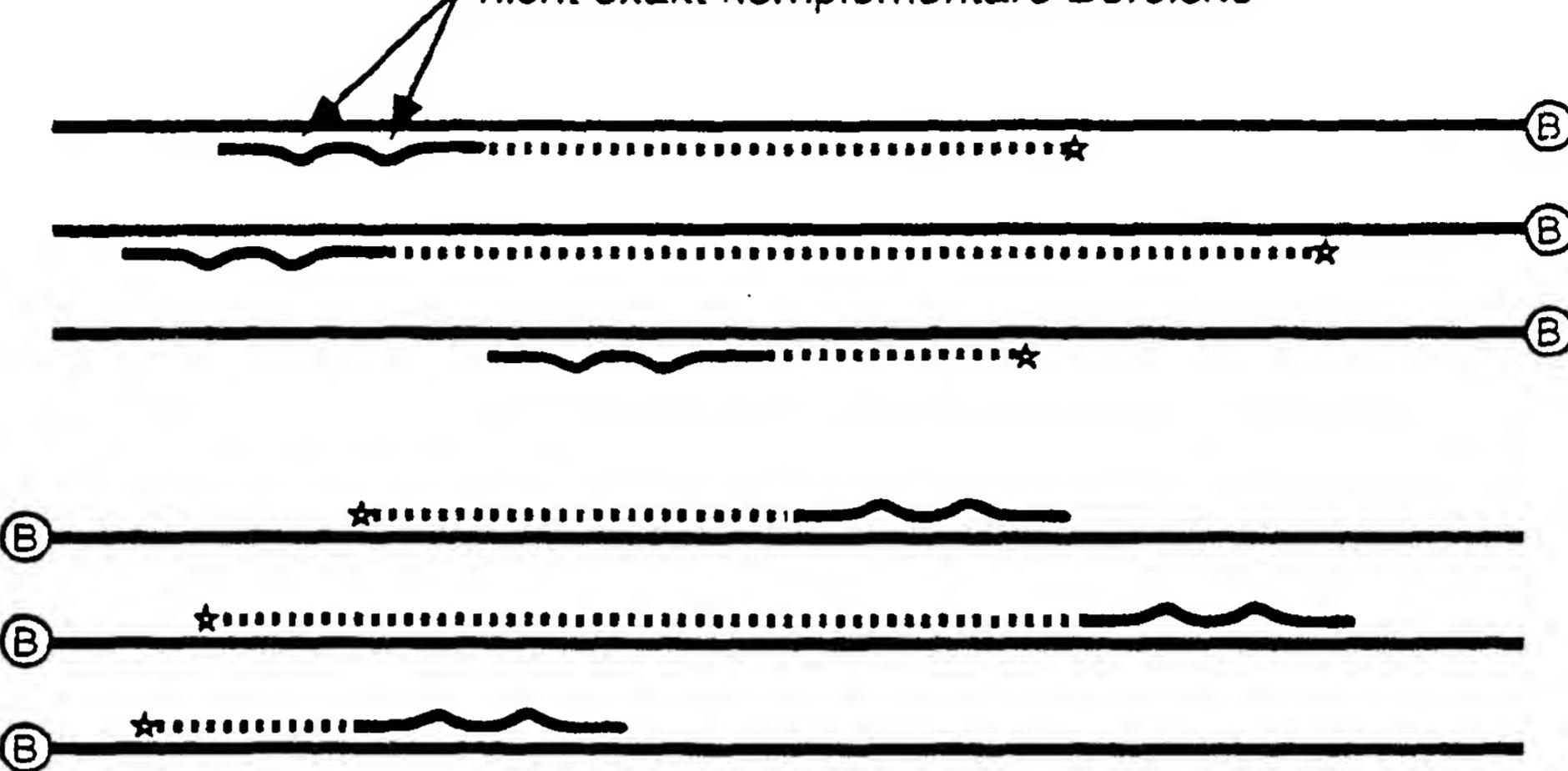


Figur 10

ERSATZBLATT (REGEL 26)

7 / 9

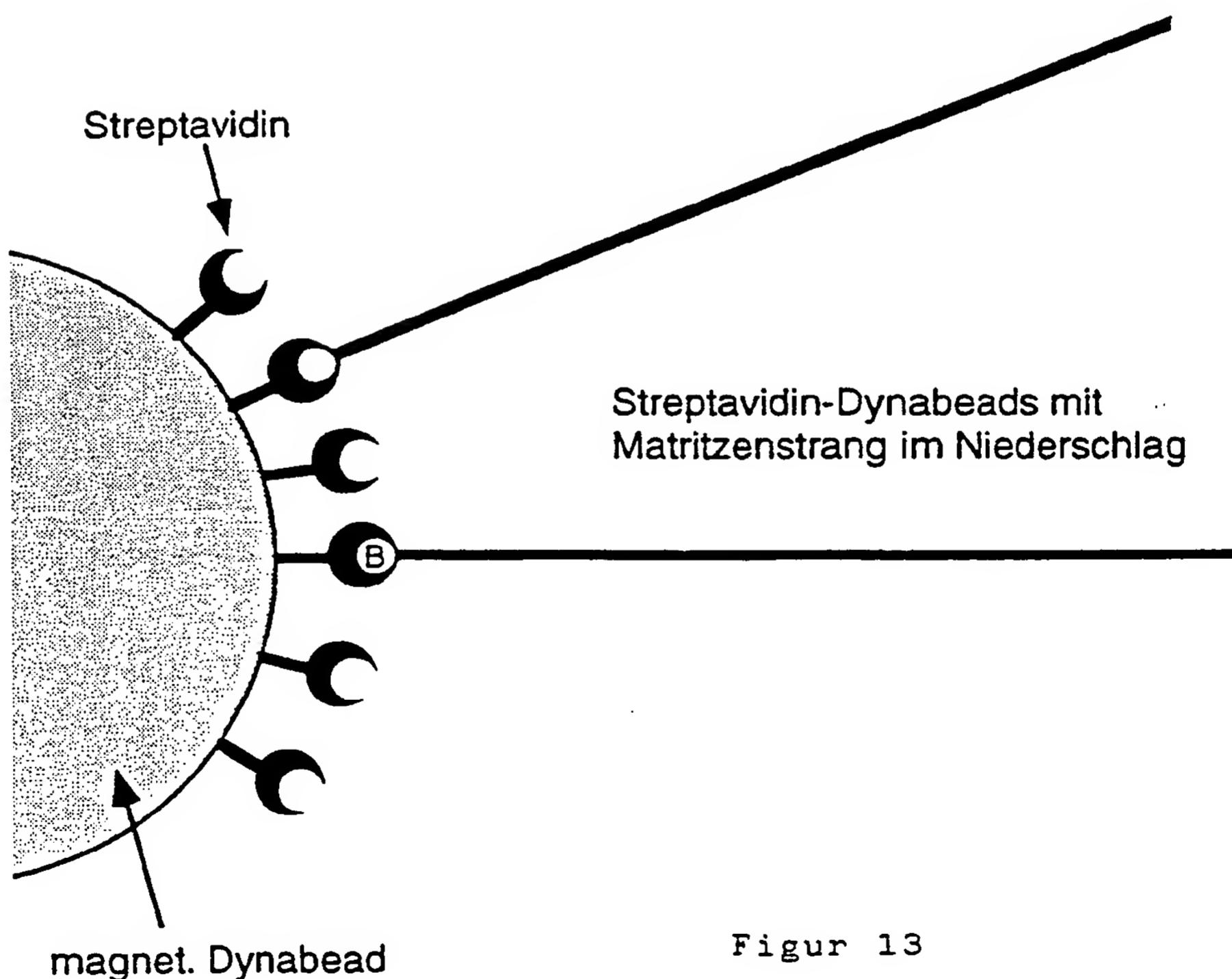
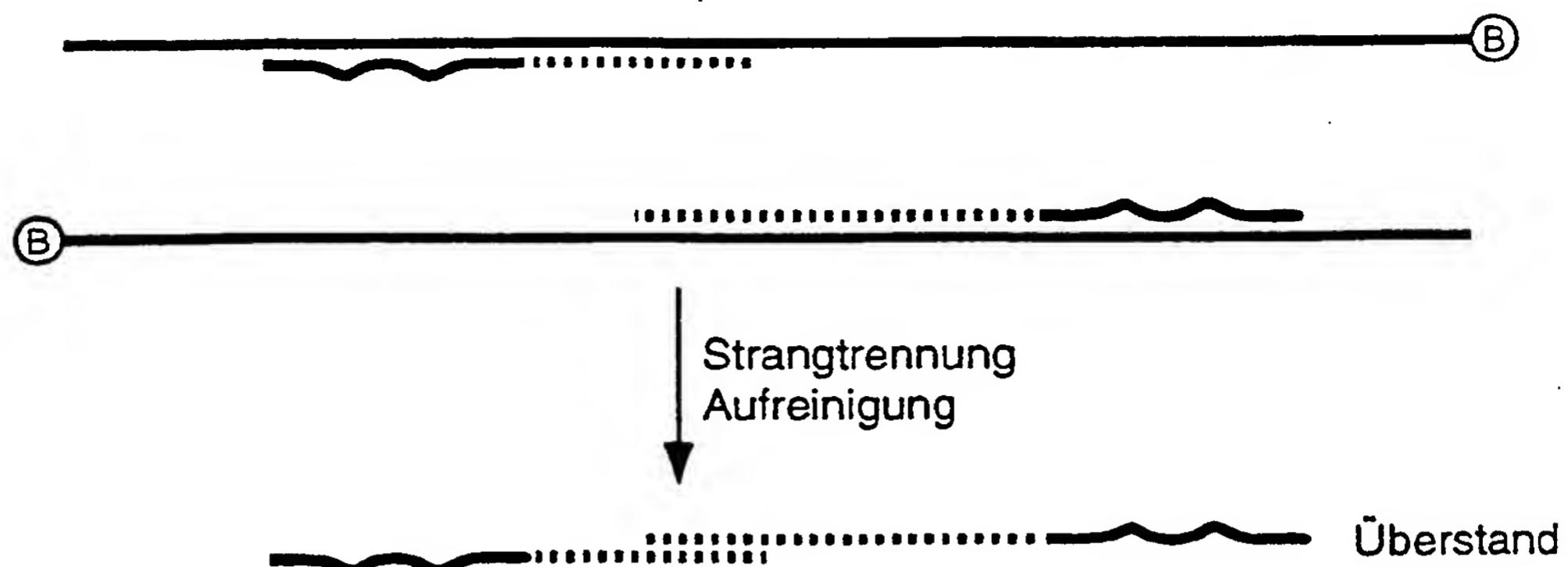
✓ nicht exakt komplementäre Bereiche



Figur 11

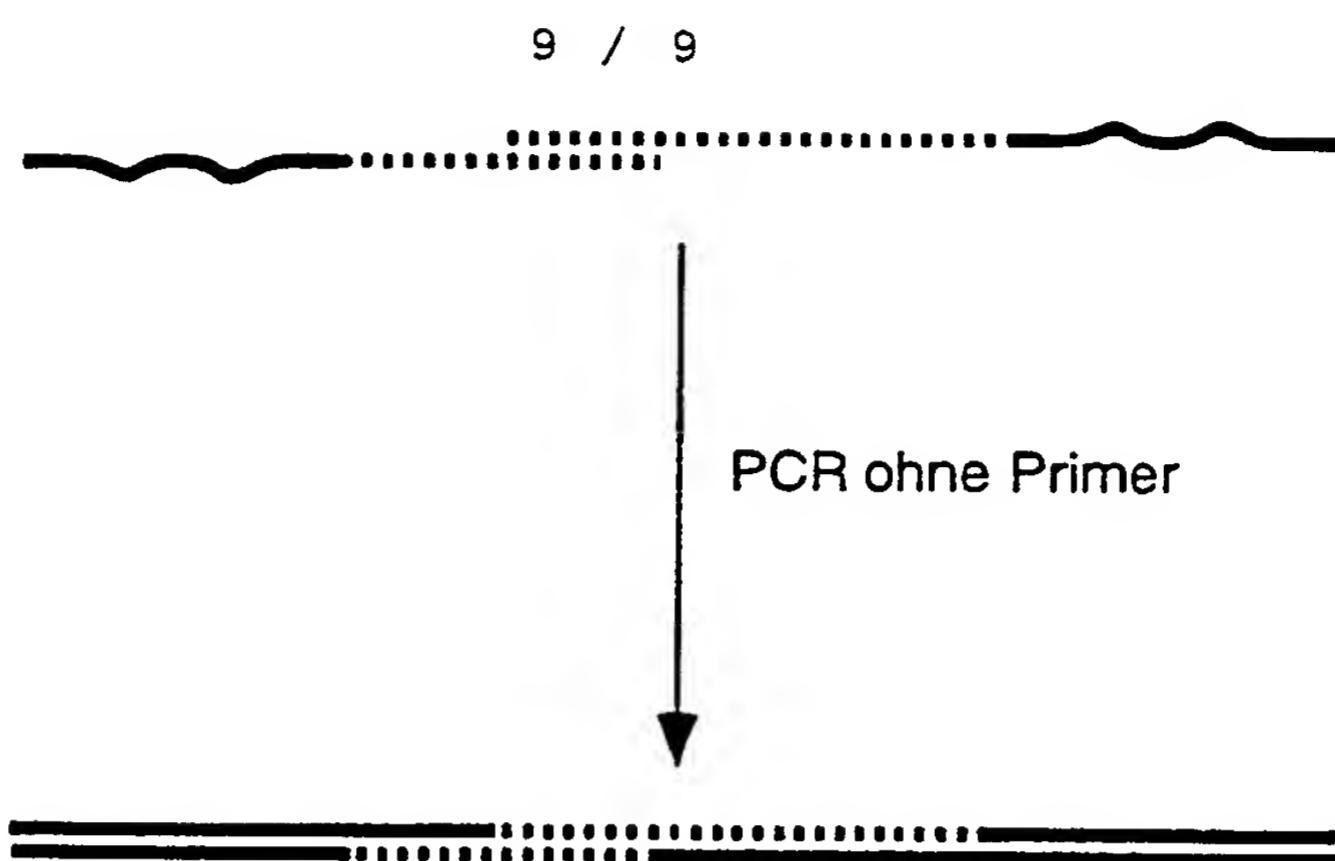
## ERSATZBLATT (REGEL 26)

8 / 9



Figur 13

ERSATZBLATT (REGEL 26)



Figur 14

ERSATZBLATT (REGEL 26)

## INTERNATIONAL SEARCH REPORT

Int. Application No

PCT/EP 94/04240

A. CLASSIFICATION OF SUBJECT MATTER  
 IPC 6 C07H21/00 C07K1/04 C12N15/10 C12P21/00

According to International Patent Classification (IPC) or to both national classification and IPC

## B. FIELDS SEARCHED

Minimum documentation searched (classification system followed by classification symbols)  
 IPC 6 C07H C07K C12N C12P

Documentation searched other than minimum documentation to the extent that such documents are included in the fields searched

Electronic data base consulted during the international search (name of data base and, where practical, search terms used)

## C. DOCUMENTS CONSIDERED TO BE RELEVANT

Category *	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
X	HELVETICA CHIMICA ACTA, vol.71, 1988, BASEL CH pages 835 - 847 MUTTER M. ET AL 'The Construction of New Proteins Part III' see page 835 - page 836 ---	1
A	WO,A,90 00626 (BAYLOR COLLEGE OF MEDICINE) 25 January 1990 see page 1, line 1 - line 15; figures ---	1
A	WO,A,92 18645 (DIAGEN INSTITUT FÜR MOLEKULARBIOLOGISCHE DIAGNOSTIK GMBH) 29 October 1992 cited in the application see the whole document -----	1

Further documents are listed in the continuation of box C.

Patent family members are listed in annex.

## \* Special categories of cited documents :

- \*A\* document defining the general state of the art which is not considered to be of particular relevance
- \*E\* earlier document but published on or after the international filing date
- \*L\* document which may throw doubts on priority claim(s) or which is cited to establish the publication date of another citation or other special reason (as specified)
- \*O\* document referring to an oral disclosure, use, exhibition or other means
- \*P\* document published prior to the international filing date but later than the priority date claimed

- \*T\* later document published after the international filing date or priority date and not in conflict with the application but cited to understand the principle or theory underlying the invention
- \*X\* document of particular relevance; the claimed invention cannot be considered novel or cannot be considered to involve an inventive step when the document is taken alone
- \*Y\* document of particular relevance; the claimed invention cannot be considered to involve an inventive step when the document is combined with one or more other such documents, such combination being obvious to a person skilled in the art.
- \*&\* document member of the same patent family

1

Date of the actual completion of the international search

27 March 1995

Date of mailing of the international search report

05.04.95

## Name and mailing address of the ISA

European Patent Office, P.B. 5818 Patentlaan 2  
 NL - 2280 HV Rijswijk  
 Tel. (+ 31-70) 340-2040, Tx. 31 651 epo nl,  
 Fax (+ 31-70) 340-3016

## Authorized officer

Day, G

**INTERNATIONAL SEARCH REPORT**

International Application No

PCT/EP 94/04240

Patent document cited in search report	Publication date	Patent family member(s)		Publication date
WO-A-9000626	25-01-90	AU-A-	3869289	05-02-90
		EP-A-	0427745	22-05-91
		JP-T-	3505966	26-12-91
-----	-----	-----	-----	-----
WO-A-9218645	29-10-92	DE-C-	4112440	22-10-92
		EP-A-	0583265	23-02-94
		JP-T-	6506355	21-07-94
-----	-----	-----	-----	-----

Form PCT/ISA/210 (patent family annex) (July 1992)

## INTERNATIONAHLER RECHERCHENBERICHT

Inn	onales Aktenzeichen
PCT/EP 94/04240	

A. KLASIFIZIERUNG DES ANMELDUNGSGEGENSTANDES				
IPK 6	C07H21/00	C07K1/04	C12N15/10	C12P21/00

Nach der Internationalen Patentklassifikation (IPK) oder nach der nationalen Klassifikation und der IPK
---

B. RECHERCHIERTE GEBIETE
--------------------------

Recherchierte Mindestprässtoff (Klassifikationssystem und Klassifikationssymbole)
IPK 6 C07H C07K C12N C12P

Recherchierte aber nicht zum Mindestprässtoff gehörende Veröffentlichungen, soweit diese unter die recherchierten Gebiete fallen
--

Während der internationalen Recherche konsultierte elektronische Datenbank (Name der Datenbank und evtl. verwendete Suchbegriffe)
---

C. ALS WESENTLICH ANGESEHENE UNTERLAGEN
---

Kategorie*	Bezeichnung der Veröffentlichung, soweit erforderlich unter Angabe der in Betracht kommenden Teile	Betr. Anspruch Nr.
X	HELVETICA CHIMICA ACTA, Bd.71, 1988, BASEL CH Seiten 835 - 847 MUTTER M. ET AL 'The Construction of New Proteins Part III' siehe Seite 835 - Seite 836 ---	1
A	WO,A,90 00626 (Baylor College of Medicine) 25. Januar 1990 siehe Seite 1, Zeile 1 - Zeile 15; Abbildungen ---	1
A	WO,A,92 18645 (DIAGEN INSITUT FÜR MOLEKULARBIOLOGISCHE DIAGNOSTIK GMBH) 29. Oktober 1992 in der Anmeldung erwähnt siehe das ganze Dokument -----	1

<input type="checkbox"/> Weitere Veröffentlichungen sind der Fortsetzung von Feld C zu entnehmen
--

<input checked="" type="checkbox"/> Siehe Anhang Patentsfamilie
---

- \* Besondere Kategorien von angegebenen Veröffentlichungen :
- \*'A' Veröffentlichung, die den allgemeinen Stand der Technik definiert, aber nicht als besonders bedeutsam anzusehen ist
- \*'E' älteres Dokument, das jedoch erst am oder nach dem internationalen Anmeldedatum veröffentlicht worden ist
- \*'L' Veröffentlichung, die geeignet ist, einen Prioritätsanspruch zweifelhaft erscheinen zu lassen, oder durch die das Veröffentlichungsdatum einer anderen im Recherchenbericht genannten Veröffentlichung belegt werden soll oder die aus einem anderen besonderen Grund angegeben ist (wie ausgeführt)
- \*'O' Veröffentlichung, die sich auf eine mündliche Offenbarung, eine Benutzung, eine Ausstellung oder andere Maßnahmen bezieht
- \*'P' Veröffentlichung, die vor dem internationalen Anmeldedatum, aber nach dem beanspruchten Prioritätsdatum veröffentlicht worden ist
- \*'T' Spätere Veröffentlichung, die nach dem internationalen Anmeldedatum oder dem Prioritätsdatum veröffentlicht worden ist und mit der Anmeldung nicht kollidiert, sondern nur zum Verständnis des der Erfindung zugrundeliegenden Prinzips oder der ihr zugrundeliegenden Theorie angegeben ist
- \*'X' Veröffentlichung von besonderer Bedeutung; die beanspruchte Erfindung kann allein aufgrund dieser Veröffentlichung nicht als neu oder auf erforderlicher Tätigkeit beruhend betrachtet werden
- \*'Y' Veröffentlichung von besonderer Bedeutung; die beanspruchte Erfindung kann nicht als auf erforderlicher Tätigkeit beruhend betrachtet werden, wenn die Veröffentlichung mit einer oder mehreren anderen Veröffentlichungen dieser Kategorie in Verbindung gebracht wird und diese Verbindung für einen Fachmann naheliegend ist
- \*'&' Veröffentlichung, die Mitglied derselben Patentsfamilie ist

1	Datum des Abschlusses der internationalen Recherche	Absendedatum des internationalen Recherchenberichts
	27. März 1995	05.04.95
	Name und Postanschrift der Internationale Recherchenbehörde Europäisches Patentamt, P.B. 5818 Patenlaan 2 NL - 2280 HV Rijswijk Tel. (+ 31-70) 340-2040, Tx. 31 651 epo nl, Fax: (+ 31-70) 340-3016	Bevollmächtigter Bediensteter Day, G

## INTERNATIONALER RECHERCHENBERICHT

Internationales Aktenzeichen

PCT/EP 94/04240

Im Recherchenbericht angeführtes Patentdokument	Datum der Veröffentlichung	Mitglied(er) der Patentfamilie		Datum der Veröffentlichung
WO-A-9000626	25-01-90	AU-A-	3869289	05-02-90
		EP-A-	0427745	22-05-91
		JP-T-	3505966	26-12-91
-----	-----	-----	-----	-----
WO-A-9218645	29-10-92	DE-C-	4112440	22-10-92
		EP-A-	0583265	23-02-94
		JP-T-	6506355	21-07-94
-----	-----	-----	-----	-----

